Defense Threat Reduction Agency
8725 John J. Kingman Road, MS
6201 Fort Belvoir, VA 22060-6201

DTRA-TR-15-55

TECHNICAL REPORT

Network Adaptability from WMD Disruption and Cascading Failures

Distribution Statement A. Approved for public release; distribution is unlimited.

April 2016

HDTRA1-10-1-0011

Biswanath Mukherjee

Prepared by:
University of California, Davis
2063 Kemper Hall
Davis, CA 95616

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 00-04-2016	**2. REPORT TYPE** Technical	**3. DATES COVERED** *(From - To)* March 15, 2010 - March 14, 2015

4. TITLE AND SUBTITLE

Network Adaptability from WMD Disruption and Cascading Failures

5a. CONTRACT NUMBER

HDTRA1-10-1-0011

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Biswanath Mukherjee

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of California, Davis
2063 Kemper Hall
Davis, CA 95616

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Threat Reduction Agency
8725 John J. Kingman Road, STOP 6201
Fort Belvoir, VA 22060-6201

10. SPONSOR/MONITOR'S ACRONYM(S)

DTRA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

DTRA-TR-15-55

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution Statement A. Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

To combat the rising risk of terrorist threats against our national interest and economic wellbeing - threats such as Weapons of Mass Destruction (WMD) attacks (which could lead to a domino effect of catastrophic failures of power, energy, telecommunications, satellite, transportation, financial, and other critical infrastructures) - this project has developed novel methods to provide protection in our information and communications networks. Main focus was to prepare the network by accounting for risk of WMD attacks in different parts of the infrastructure, and recover from them. These techniques can be applied to military infrastructure to counteract multiple, correlated, and cascading failures induced by large-scale attacks.

15. SUBJECT TERMS

Network adaptability; WMD disruptions; Cascading failures, Survivability in cloud; Degraded service

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Paul Tandy
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	70	**19b. TELEPHONE NUMBER** *(include area code)* 703-767-4663

UNIT CONVERSION TABLE

U.S. customary units to and from international units of measurement[*]

U.S. Customary Units	Multiply by → ← Divide by[†]		International Units
Length/Area/Volume			
inch (in)	2.54	$\times 10^{-2}$	meter (m)
foot (ft)	3.048	$\times 10^{-1}$	meter (m)
yard (yd)	9.144	$\times 10^{-1}$	meter (m)
mile (mi, international)	1.609 344	$\times 10^{3}$	meter (m)
mile (nmi, nautical, U.S.)	1.852	$\times 10^{3}$	meter (m)
barn (b)	1	$\times 10^{-28}$	square meter (m^{2})
gallon (gal, U.S. liquid)	3.785 412	$\times 10^{-3}$	cubic meter (m^{3})
cubic foot (ft^{3})	2.831 685	$\times 10^{-2}$	cubic meter (m^{3})
Mass/Density			
pound (lb)	4.535 924	$\times 10^{-1}$	kilogram (kg)
unified atomic mass unit (amu)	1.660 539	$\times 10^{-27}$	kilogram (kg)
pound-mass per cubic foot (lb ft^{-3})	1.601 846	$\times 10^{1}$	kilogram per cubic meter (kg m^{-3})
pound-force (lbf avoirdupois)	4.448 222		newton (N)
Energy/Work/Power			
electron volt (eV)	1.602 177	$\times 10^{-19}$	joule (J)
erg	1	$\times 10^{-7}$	joule (J)
kiloton (kt) (TNT equivalent)	4.184	$\times 10^{12}$	joule (J)
British thermal unit (Btu) (thermochemical)	1.054 350	$\times 10^{3}$	joule (J)
foot-pound-force (ft lbf)	1.355 818		joule (J)
calorie (cal) (thermochemical)	4.184		joule (J)
Pressure			
atmosphere (atm)	1.013 250	$\times 10^{5}$	pascal (Pa)
pound force per square inch (psi)	6.984 757	$\times 10^{3}$	pascal (Pa)
Temperature			
degree Fahrenheit (^{o}F)	[T(^{o}F) − 32]/1.8		degree Celsius (^{o}C)
degree Fahrenheit (^{o}F)	[T(^{o}F) + 459.67]/1.8		kelvin (K)
Radiation			
curie (Ci) [activity of radionuclides]	3.7	$\times 10^{10}$	per second (s^{-1}) [becquerel (Bq)]
roentgen (R) [air exposure]	2.579 760	$\times 10^{-4}$	coulomb per kilogram (C kg^{-1})
rad [absorbed dose]	1	$\times 10^{-2}$	joule per kilogram (J kg^{-1}) [gray (Gy)]
rem [equivalent and effective dose]	1	$\times 10^{-2}$	joule per kilogram (J kg^{-1}) [sievert (Sv)]

[*] Specific details regarding the implementation of SI units may be viewed at http://www.bipm.org/en/si/.

[†] Multiply the U.S. customary unit by the factor to get the international unit. Divide the international unit by the factor to get the U.S. customary unit.

2015-11-16

Table of Contents

This final report summarizes all of the research activities on our DTRA project entitled *"Network Adaptability from WMD Disruption and Cascading Failures,"* and conducted under the supervision of PI Professor Biswanath Mukherjee at the University of California, Davis (UCD). The period of performance for this project was between March 15, 2010, and March 14, 2015.

1. What are the major goals of the project?

This project investigated novel methods to provide protection against WMD stressors in mesh networks: by means of connection re-provisioning, multi-path routing, and data replication. These techniques can be applied to military infrastructure to counteract multiple, correlated, and cascading failures induced by WMD stressors.

We developed novel paradigms for both network and cloud survivability with multi-path routing, re-provisioning, network mapping, data-replication, and content protection algorithms. Also, we have considered newer trends – such as virtualization and cloud computing – in government, military, and commercial networks, e.g., the White House's "Cloud-First" policy, in which IT services of Federal Agencies are being moved to the cloud gradually [1].

Our research contributes to the science of network survivability, and yields fundamental insights on the design, operation, and performance of vital military network backbone and cloud infrastructures subject to WMD stressors. In particular, our goal was to ensure WMD-resilient networking in three phases:

1. Normal WMD Preparedness (by accounting for risk of attacks in different parts of the infrastructure);
2. Enhanced WMD Preparedness (under more-accurate intelligence on potential attacks); and
3. Post-WMD-Attack network and service survivability.

All tasks of the project are presented below.

1: Characterizing the impact of WMD stressors in WDM networks

2: WMD-aware re-provisioning algorithms

3: Analysis of WMD survivability provided by multipath routing

4: Degraded-service multipath scheme against WMD attacks

5: Multi-state multipath provisioning to combat node failures

6: Network protection architectures based on data replication

7: Provisioning of highly-critical services with data replication

8: WMD-risk-aware virtual-network mapping and re-mapping

9: Virtual network mapping for content connectivity in cloud

10: WMD-aware datacenter placement and data replication

11: Adaptation to change of risk level for cloud services

12: Service recovery after a WMD attack

2. What was accomplished under these goals?

Optical backbone networks employ mesh structures to provide high-bandwidth connectivity across large distances, and they are integral to our national security and economic well-being. These backbone networks work as a scalable substrate to multiplex higher-layer networks (e.g., SONET, Ethernet, IP/MPLS, etc.) Wavelength-Division Multiplexing (WDM) is the most popular technology for optical networks due to its flexibility and robustness. Given the high complexity and scalability of optical WDM backbone networks, and the dependency of other networks on them, the vulnerability of optical backbone networks to large-scale failures induced by WMD stressors is a major concern. Various techniques exist to provide fast protection at the optical layer, but they are developed for limited faults, such as a single fiber cut, but not for the potentially-large extent of WMD attacks [2]. Also protection in cloud against WMD attacks has not been studied in detail [3].

Network adaptability from WMD disruption and cascading failures requires advanced protection methods, exploiting the knowledge of risky regions and forecasted attacks. The events during the aftermath of an attack should also be considered. For instance, methods to re-arrange network resources and services on a partially-damaged network, i.e., properties of a *self-organized network*, should be developed; and new algorithms to manage post-attack traffic deluge and to relieve the rescue operations after an attack, with the knowledge of the post-attack failures, should be investigated. Since cloud services today are an integral part of our society and massive amount of content/services have been created and shared over the cloud, loss/disruption of critical content/services caused by WMD attacks can significantly affect the socio-economy and security of our nation.

We investigated the possible target zones in the network; risk-aware routing to reduce network disruption in the event of a WMD attack; multi-path provisioning to provide partial protection when there is not enough capacity to provide full protection, e.g., after an attack; the nature of possible disruption in upper-layer networks (e.g., virtual network, IP network, etc.) caused by disruption in the optical layer (physical network) induced by a WMD attack; and methods to provide protection in upper-layer networks. Moreover, for survivability in the cloud, we developed WMD-risk aware virtual-network mapping and content-placement schemes to ensure content connectivity in the cloud, and protection schemes exploiting degraded-service tolerance of services and manycasting. A technical summary of the accomplishments for all tasks of the project is presented under three main sections below; namely: A. Characterizing the impact of WMD stressors in WDM networks, B. WMD-Risk-Aware Provisioning and Recovery Methods, and C. WMD Survivability in Cloud Networks.

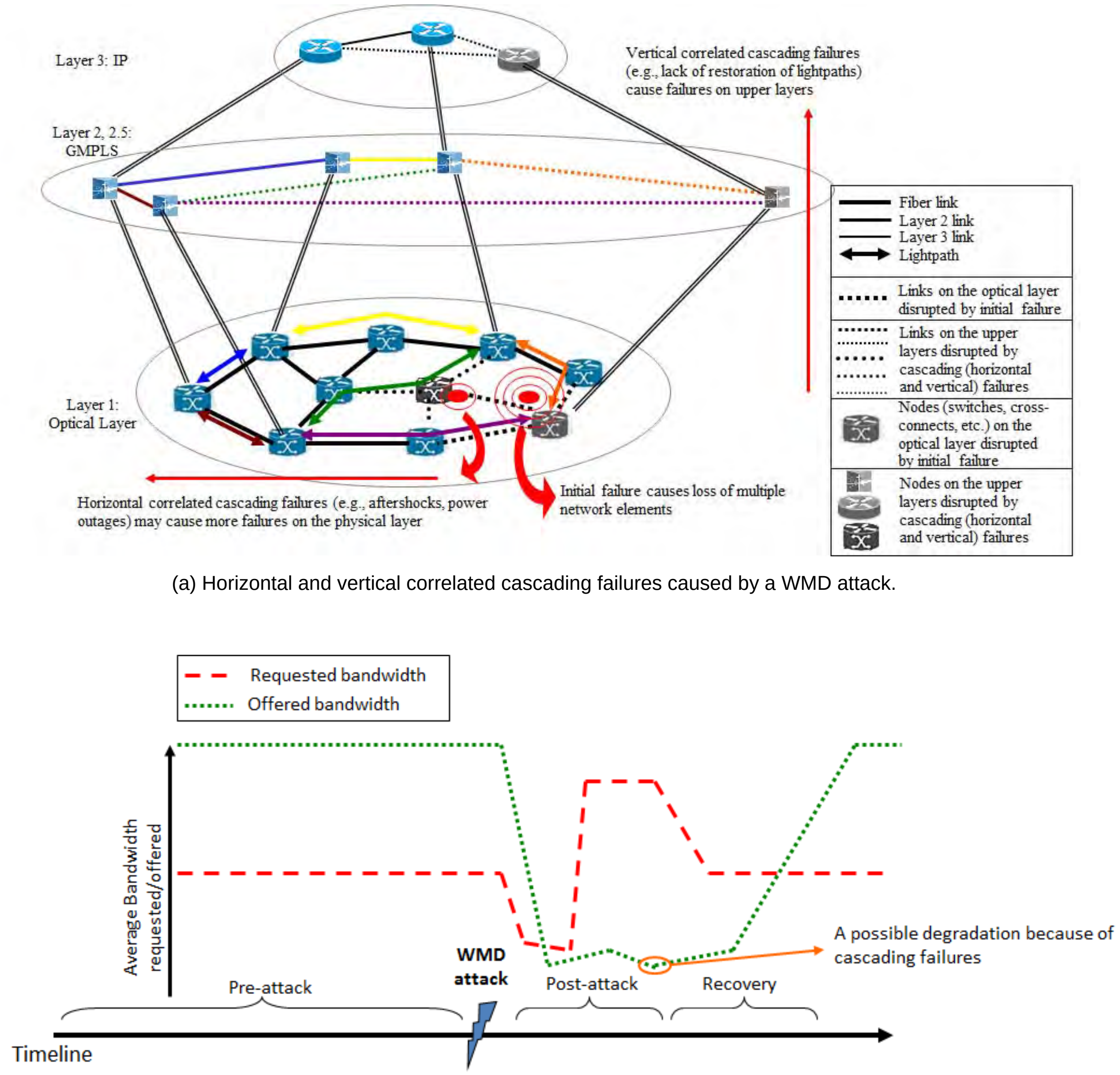
It is also important to quantify, wherever possible, the performance improvements of our approaches relative to prior art. These quantitative improvements are summarized at the beginning of the section on Impact (Section 4). These improvements can be appreciated only after our methods are described, so the quantitative improvements are summarized towards the end of this report.

A. Characterizing the impact of WMD stressors in WDM networks

Characteristics of WMD Disruptions

Our project investigated the characteristics of WMD attacks and the nature of possible disruption in backbone networks. A WMD attack may cause a set of multiple correlated cascading failures. The initial shock, when a WMD attack occurs in a region, typically may cause multiple network-element failures at the physical (optical) layer. But, after the initial failure, some correlated incidents (e.g., secondary attacks, power outage) might cause other failures at the optical layer (which we call *horizontal correlated cascading failures*). The optical layer provides lightpaths (i.e., wavelength channels/circuits) to form connectivity between nodes (switches, routers, etc.) and datacenters of the upper layers (e.g., virtual networks, IP, SONET, etc.), so lack of restoration of lightpaths can cause failures at upper layers and data loss (which we call *vertical correlated cascading failures*). Figure 1(a) shows an example of a WMD attack inducing horizontal and vertical correlated cascading failures.

The available network capacity and the bandwidth demand also vary before, during, and after an attack. Figure 1(b) shows a typical timeline before and after an attack with an example of average offered and requested bandwidth utilization. Telecom networks usually have some unused capacity, called excess capacity (EC), to accommodate traffic fluctuations and to avoid capacity exhaustion. During a WMD attack, some military facilities supported by backbone networks may be temporarily closed, and some of the network resources may suffer major failures which may decrease both offered and requested bandwidth. Typically, there will be many inquiries to/from the affected area after an attack, and this will cause a traffic flood which may cause blocking or congestion of services required for rescue operations. After the post-attack period, when the network equipment is recovered, intelligent relief techniques can restore services while increasing the offered bandwidth gradually.


Layer 3: IP
Layer 2, 2.5: GMPLS
Layer 1: Optical Layer
Vertical correlated cascading failures (e.g., lack of restoration of lightpaths) cause failures on upper layers
Horizontal correlated cascading failures (e.g., aftershocks, power outages) may cause more failures on the physical layer
Initial failure causes loss of multiple network elements
Fiber link
Layer 2 link
Layer 3 link
Lightpath
Links on the optical layer disrupted by initial failure
Links on the upper layers disrupted by cascading (horizontal and vertical) failures
Nodes (switches, cross-connects, etc.) on the optical layer disrupted by initial failure
Nodes on the upper layers disrupted by cascading (horizontal and vertical) failures
Requested bandwidth
Offered bandwidth
Average Bandwidth requested/offered
WMD attack
Pre-attack
Post-attack
Recovery
A possible degradation because of cascading failures
Timeline

(a) Horizontal and vertical correlated cascading failures caused by a WMD attack.

(b) Requested and offered bandwidth before and after a WMD attack.

Figure 1. Nature of WMD disruptions in backbone networks.

Determination of the Geographical Locations of a Network Topology

To prepare a network against WMD attacks, we need to determine the possible risky zones in the network and the attacks' possible impacts on the network. Since WMD attacks cause failures on vertices and edges, the exact locations of the components of a network are very significant. The exact locations of each vertex can be easily discovered. *However, the physical links between vertices are not straight lines, but they can be approximately estimated by other known structures.* For instance, the assumption “communication lines are close to highways and/or railroads” is very reasonable, because any point on a link should be accessible in case of a failure. Thus, we have constructed a model of a sample US nation-wide backbone optical network with physical locations of all vertices and links by matching US-wide topology with US transportation map. WMD attacks usually target populated cities and cities where important government, military, or resource facilities are located. As an example target set, we define the 11 most populated cities and Washington DC as possible WMD targets, and we consider the probability that a city is targeted to be proportional to its population or its importance. The physical topology and example target set are shown in Fig. 2. The numbers show the probability of an attack where p is the probability of a WMD attack to US.

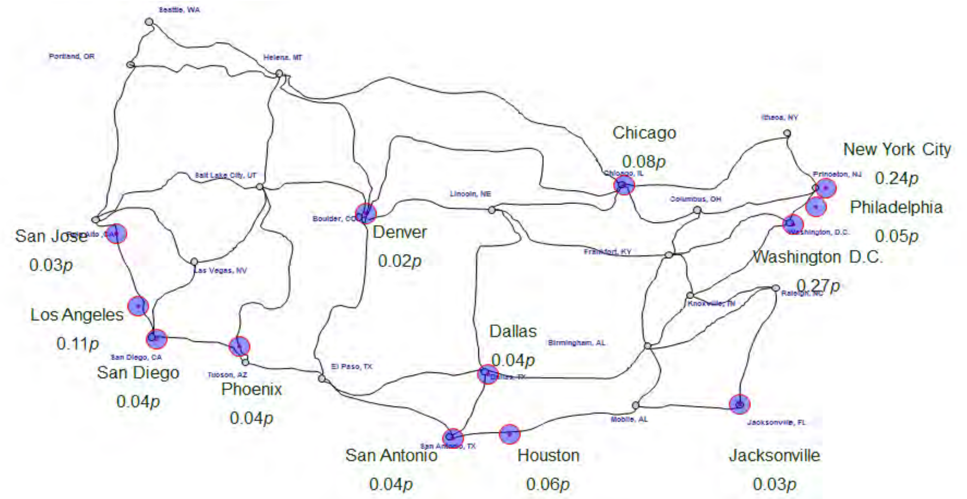


Figure 2. Physical topology with example target set.

Risk Analysis

Providing full protection for all services against WMD disruption with traditional approaches may require massive and economically-unsustainable bandwidth overprovisioning and data storage, as some attacks are difficult to predict, statistically rare, and may create large-scale failures. To combat WMD attacks, we need a paradigm shift from 100% deterministic recovery of a specific type of failure (e.g., single link, double link, shared risk group, etc.) to a probabilistic approach which can minimize the risk of outages/data loss and can possibly provide full protection for mission-critical services and reduced services (degraded services) for others, even in the event of a WMD disruption.

For this paradigm shift, we define a risk parameter by exploiting possible set of targets information, capturing following questions:

- What can happen? (e.g., an attack on a possible target)
- How likely that it will happen? (e.g., probability of an attack)
- If it happens, what are the consequences? (e.g., loss after an attack)

WMD attacks’ scales and damage zones (usually defined within a certain radius) give us which network elements might be disconnected. Each WMD attack n in possible attack set (N) can be represented by S_n ($|S_n| = |E|$), set of binary valuables s^{ij}_n (which is 1 if the fiber link (i, j) is disconnected by disaster n, and 0 otherwise[4]), and p_n^d as the probability that attack n causes failures of network elements.

These failures may cause not only connection loss, but also data loss and disruption/disconnection of virtual networks. We define a penalty value to understand the loss of network operator in case of an attack. Below example shows how we define risk for connection loss.

Connections' priorities/requirements are usually different where losing some of the connections might cause more problems because of their importance than others. Let c_t and R^n_t denote the penalty per unit time of connection t and binary variable which is equal to 1 if connection *t* is lost with attack *n*, respectively. The loss (per unit time) in case of attack *n* will be $\sum_{t \in T} c_t R_t^n$, where *T* is the set of connections. Thus, a risk model is defined as:

$$\sum_{n \in N} \left(\sum_{t \in T} c_t R_t^n \right) p_d^n \qquad (1)$$

Network Preparedness and Adaptability against WMD Disruption

Since WMD attack failures may lead to multiple correlated cascading failures, the protection before an attack and restoration after an attack require careful planning and have different dimensions (e.g., lightpath protection, service protection, facility protection, etc.). By exploiting the risk information found through risk analysis, precautions can be taken before an attack by avoiding risky zones and/or backing up lightpaths/services. After an attack is predicted, the network can be better prepared by reorganizing resources and by re-disseminating data. Even if risky regions are avoided or are adequately protected, failures might still cause network disruptions. Unless the network operator is willing to invest very high capital expenditure for massive overprovisioning of the network, WMD-aware precautions can only minimize possible network disruptions, but cannot eliminate them. Thus, restoration of services is required after a WMD disruption.

Following the timeline of WMD disruption events as shown in Fig. 1(b), we consider WMD-resilient networking in three phases:

1. **Normal Preparedness:** Network operator should proactively take necessary actions to minimize the network disruptions and data loss in case of a WMD attack. In this phase, knowledge of possible risky regions (i.e., risk information) would help to utilize network resources and disseminate data accordingly. Note that, as failures due to WMD attacks are large and geographically collocated, recovery in upper (virtual, IP, etc.) layers requires cross-layer signaling between the optical (physical) layer and the upper layers such that the upper layer can utilize the geographic and failure information of the physical-layer equipment. Cross-layer signaling is a challenging problem. Also as a single failure propagates from a lower layer to an upper layer, causing multiple failures, it may require more resources and time to recover. To enable faster, easier, and scalable automatic response to large-scale failures due to WMD attacks, protection schemes have been developed mainly for the optical layer.
2. **Enhanced Preparedness:** If the time and location of an attack (through intelligence information) is predicted, the network operator can respond by re-allocating network resources and re-disseminating data, and possibly by relocating hardware resources.

   If a WMD attack is predicted through intelligence information, the risk of network disruption can be minimized by re-arranging network resources such that network elements that fall in possible attack region are used as less as possible. In such scenarios, reactive measures can be taken to adapt to the changing risk levels of attacks and to enhance the protection of the network. Risk-aware provisioning can use the changing risk inputs. Data/content protection can also be provided by replicating data/content from a datacenter under high risk in case of an upcoming WMD attack to a safe location. These pre-WMD-attack actions, namely re-allocating network resources and replicating data, should be done considering cascading effects of the attack (e.g., power outage).

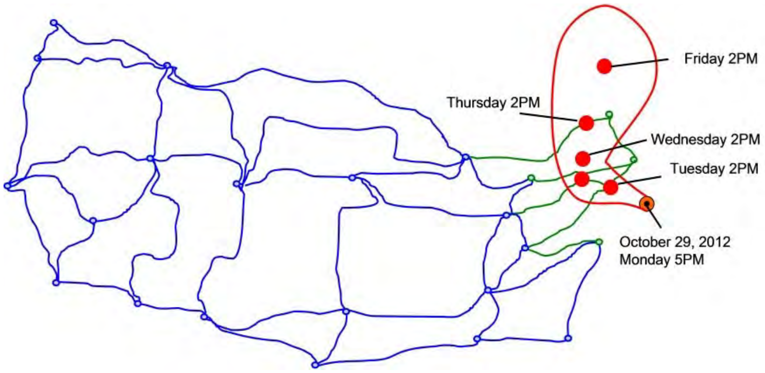


Figure 3. Path of Hurricane Sandy projected on October 29, 2012, 5 PM (red) and network elements that might be affected by the hurricane and/or its correlated cascading failures (green).

Similar techniques can be applied to provide protection against a forecasted natural disaster. We explain the concept of enhanced preparedness using the example of Hurricane Sandy. Figure 3 shows the projection of Hurricane Sandy's path on October 29, 2012, with red lines, and network elements that might be affected by the hurricane and/or its correlated cascading failures with green lines and circles. Any data in any datacenter which is located in these green circles should be replicated to safe locations to avoid the loss of data due to failures by the hurricane. The connections traversing the links and nodes shown in green should also be reprovisioned by exploiting the EC on links shown in blue.

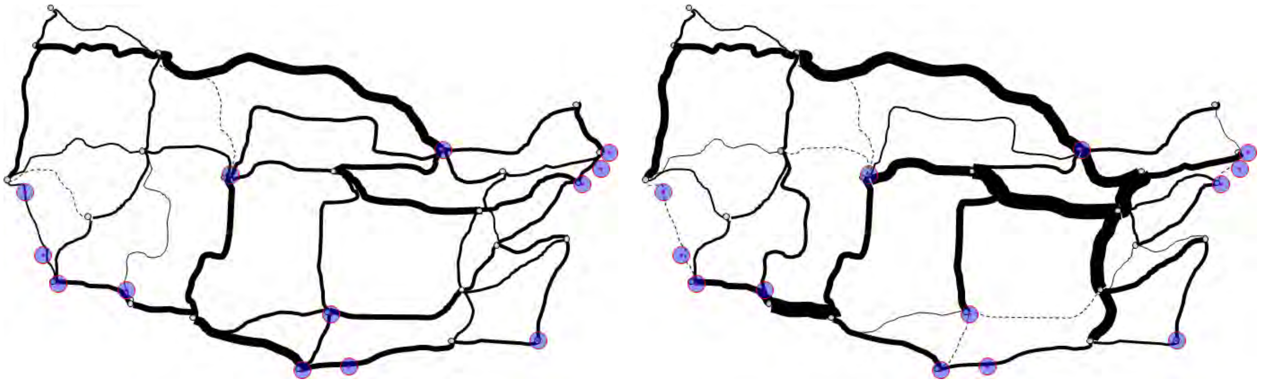
3. **Post-WMD-Attack Actions:** After an attack, network resources can become limited, so services should be provided with as much bandwidth as possible (degraded services). Disrupted connections and virtual networks can be reprovisioned on the surviving network resources. During the reprovisioning phase, cascading failures (both horizontal and vertical) should also be taken into consideration. An important feature of these cascading failures is that they tend to be more predictable from the damage and location of the initial failure, and this prediction can be used to reorganize network connectivity in view of such possible cascading failures.

By adopting these techniques, a network is necessarily transformed into a *self-organized network* which can adapt itself to changing conditions due to WMD disruptions. We elaborate our research of 1) network survivability, and 2) cloud survivability by considering these three counteracts against WMD attacks in the following two sections.

B. WMD-risk-aware provisioning and recovery methods

A Prevention Method: WMD-Aware Provisioning

A provisioning approach which minimizes the risk will reduce the loss in case of an attack. We formulated risk-aware provisioning methods (Integer Linear Program and heuristic algorithms) which choose less-risky regions for connections during provisioning (i.e., normal preparedness) to reduce the loss in case of attacks. The details of the risk model and the provisioning methods have been reported in [4]. Figure 4 shows how risk-aware provisioning uses links which traverse less-risky regions more than others (the thicknesses of the links are proportional to the resource consumption on the links) for the risky zones shown in Fig. 2.



(a) Risk-unaware provisioning.

(b) Risk-aware provisioning.

Figure 4. Comparison between (a) risk-unaware and (b) risk-aware provisioning (thicknesses of the links are proportional to their bandwidth utilization). Note the higher utilization of links in less-risky regions in risk-aware provisioning (target zones are shown in circles).

This approach encourages the network operator to choose less-risky region during reprovisioning. We formulate the risk-aware provisioning problem into an Integer Linear Program (ILP) as follows.

Given:

- $G(V, E)$: network topology.
- $T = \{t = < s_t, d_t, c_t >\}$: Set of connections to be reprovisioned with source (s_t), destination (d_t), and cost of losing connection (c_t).
- W_{ij} : number of wavelengths on link (i, j).
- $N = \{n = < S_n, p_n^d >\}$: set of attacks with S_n ($|S_n|=|E|$) is the set of binary values s^{ij}_n (which is 1 if link (i, j) is disconnected in case of event n, and 0 otherwise), and p_n^d, the probability that attack n causes failures of network elements.

Binary variables:

- R^n_t : 1 if connection t is lost in case of attack n otherwise 0.
- R^t_{ij} : 1 if connection t is routed on fiber link (i, j), otherwise 0.

Objective:

$$\min\left(\sum_{n\in N}\left(\sum_{t\in T}c_t R_t^n\right)p_d^n\right)+\varepsilon\sum_{t\in T}\sum_{(i,j)\in E}R_{ij}^t \qquad (2)$$

Binarization:

$$R_t^n \geq \frac{1}{M}\sum_{(i,j)\in E}s_n^{ij}R_{ij}^t, \quad \forall t\in T, n\in N \qquad (3a)$$

$$R_t^n \leq \sum_{(i,j)\in E}s_n^{ij}R_{ij}^t, \quad \forall t\in T, n\in N \qquad (3b)$$

Flow-conservation constraint:

$$\sum_{(i,s)\in E}R_{is}^t - \sum_{(s,j)\in E}R_{sj}^t = -1, \quad \forall t\in T \qquad (4a)$$

$$\sum_{(i,d)\in E} R_{id}^{t} - \sum_{(d,j)\in E} R_{dj}^{t} = 1, \quad \forall t \in T \tag{4b}$$

$$\sum_{(k,j)\in E} R_{kj}^{t} - \sum_{(j,k)\in E} R_{jk}^{t} = 0, \quad \forall t \in T, k \neq s, d \tag{4c}$$

Link-capacity constraint:

$$\sum_{t\in T} R_{ij}^{t} \leq W_{ij}, \quad \forall (i,j) \in E \tag{5}$$

In the objective function (Eq. (2)), the first term is risk and the second term is required to avoid unnecessarily long paths. *ε* is a small number (e.g., 10^{-5}) to keep the risk minimization as the main objective. Eq. (3) is required to understand if a connection is lost by a disaster, where *M* is a large number (e.g., 10^{3}). Eqs. (4) and (5) show flow-conservation and link-capacity constraints. Since ILP returns an optimal solution, but has high time complexity, we developed a heuristic for the problem, shown below. Eq. (6) in Step 3 encourages connections to be provisioned on links which are not in any disaster zone. If this is inevitable, then links which traverse less-risky regions are preferred.

Algorithm 1 Risk-Aware Provisioning

```
1. Create set of connections to be provisioned (i.e., $T_p = T$) and sort them in descending order with respect to their $c_t$ values.
2. Take first connection $t$ from $T_p$.
3. Update link cost by following equation:
```

$$C(e) = \begin{cases} \infty & if\ F_{ij} = 0, \\ \varepsilon & if\ \forall n \in N : s_n^{ij} = 0, \\ a_{ij} + \varepsilon \times (W_{ij} - F_{ij}) & o.w. \end{cases} \tag{6}$$

```
   where $F_{ij}$ is the number of free wavelengths on link $(i, j)$ and $a_{ij} = (1 - \max_{n\in N} p_d^n s_n^{ij})$.
4. Find shortest path from $s_t$ to $d_t$ and provision $t$ on this path.
5. Remove $t$ from $T_p$.
6. if $|T_p| \neq 0$, then go to Step 2, else terminate.
```

Network operators usually protect connections against single link failures by provisioning connections on a primary path and a link-disjoint dedicated backup path (i.e., dedicated path protection (DPP)). If DPP exists, connections will be lost in case of a disaster, if both primary and backup paths are disconnected. We can modify our model by introducing following binary variables.

- B^{n}_{t} : 1 if backup path of connection t is lost with attack *n* otherwise 0.
- B^{t}_{ij} : 1 if backup path of connection t is routed on fiber link (*i*, *j*), otherwise 0.
- Z^{n}_{t} : 1 if both primary and backup path of connection t is lost with attack *n* otherwise 0.

The new objective function is shown below:

$$\min\left(\sum_{n\in N}\left(\sum_{t\in T} c_t Z_t^{n}\right) p_d^{n}\right) + \varepsilon \sum_{t\in T} \sum_{(i,j)\in E} \left(R_{ij}^{t} + B_{ij}^{t}\right) \tag{7}$$

Similar to Eqs. (3) and (4), binarization and flow-conservation constraints for backup path are shown below:

$$B_t^n \geq \frac{1}{M} \sum_{(i,j)\in E} s_n^{ij} B_{ij}^t, \quad \forall t \in T, n \in N \qquad (8a)$$

$$B_t^n \leq \sum_{(i,j)\in E} s_n^{ij} B_{ij}^t, \quad \forall t \in T, n \in N \qquad (8b)$$

$$\sum_{(i,s)\in E} B_{is}^t - \sum_{(s,j)\in E} B_{sj}^t = -1, \quad \forall t \in T \qquad (9a)$$

$$\sum_{(i,d)\in E} B_{id}^t - \sum_{(d,j)\in E} B_{dj}^t = 1, \quad \forall t \in T \qquad (9b)$$

$$\sum_{(k,j)\in E} B_{kj}^t - \sum_{(j,k)\in E} B_{jk}^t = 0, \quad \forall t \in T, k \neq s, d \qquad (9c)$$

Following equations are required to understand if both primary and backup paths of a connection are disconnected by an attack:

$$Z \leq R_t^n, \quad \forall t \in T, n \in N \qquad (10a)$$

$$Z \leq B_t^n, \quad \forall t \in T, n \in N \qquad (10b)$$

$$Z \geq R_t^n + B_t^n - 1, \quad \forall t \in T, n \in N \qquad (10c)$$

The equation to ensure that primary and backup paths are link-disjoint is shown below:

$$R_{ij}^t + B_{ij}^t + R_{ji}^t + B_{ji}^t \leq 1, \quad \forall t \in T, (i,j) \in E \qquad (11)$$

Link-capacity constraint is also changed as follows:

$$\sum_{t\in T} \left(R_{ij}^t + B_{ij}^t \right) \leq W_{ij}, \quad \forall (i,j) \in E \qquad (12)$$

For a heuristic solution, we can use Alg. 1 to provision primary paths. After provisioning the primary path, we can find a backup path by updating link costs with the following equation:

$$C(e) = \begin{cases} \infty & if\ F_{ij} = 0 \vee (i,j) \in r_t, \\ \varepsilon & if\ (i,j) \notin D_t, \\ a_{ij} + \varepsilon \times (W_{ij} - F_{ij}) & o.w. \end{cases} \qquad (13)$$

where r_t and D_t are the set of links on primary path and those which traverses a disaster zone, respectively. Therefore, paths that do not share any risk regions are preferable. But, if this is infeasible, the links traversing less risky regions are chosen.

Illustrative Numerical Examples: ILP vs. Heuristic

We consider a small 10-node topology shown in Fig. 5 with two possible WMD targets where probabilities of attacks are 0.2*p* and 0.5*p*. The network has 16 wavelengths per link in each direction and wavelength conversion (e.g., optical switches with optical-electronic-optical (O/E/O) conversion). Traffic is uniformly distributed among node pairs. We conduct numerical examples for single path (SP) solution and dedicated path protection (DPP), where number of connections $|T|$ varies between 10 and 55.

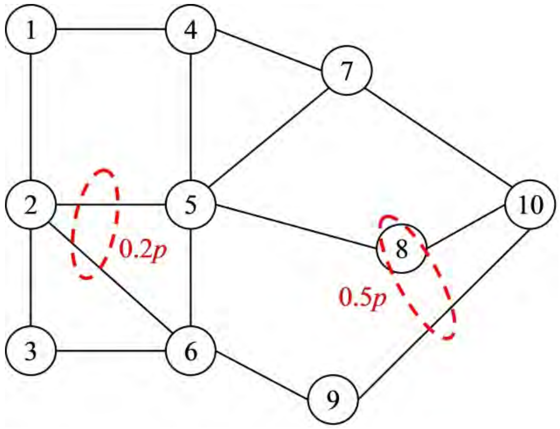


Figure 5. Sample 10-node topology.

Figure 6 shows average risk reduction results for 50 runs (compared to an optimization approach which minimizes resource use). Note that, for DPP, when $|T| \geq 35$, there is no feasible solution. For both SP and DPP approach, heuristic shows very close performance to the optimal solution. Thus, we provide numerical examples with heuristic approaches by using the US topology and WMD targets shown in Fig. 2.

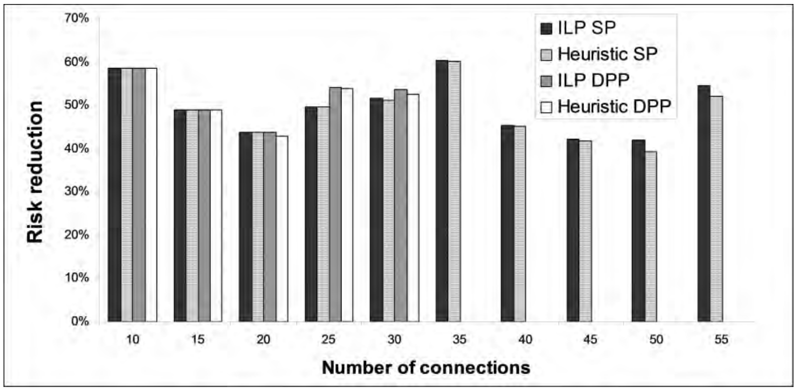


Figure 6. Risk reduction for ILP and heuristic solutions.

Numerical Examples for US-Wide Topology

We obtain the traffic demand such that connection requests are proportional to the populations of source and destination. We assign connections different c_t (penalty per unit time) values: 10α, 5α, 3α, and α with the distribution 1: 2: 3: 4, where the number of connections with high penalty is less than the others. α is some value to parameterize the penalty (e.g., penalty per hour). The number of generated connections varies between 50 and 450 for single-path (SP) provisioning and between 50 and 200 for DPP. Results shown below are average of 50 generations with 95% confidence interval.

We evaluated our approaches in terms of risk and penalty reduction and increase in (consumed) resources compared to an approach which minimizes resource consumption. Figure 7 shows risk reduction (marked with triangles) and increase in resources (marked by squares) for SP (shown by solid lines) and DPP (shown by dashed lines) scenarios. Risk reduction decreases with increasing traffic load, because more connection paths traverse disaster zones. Risk reduction is between 25% and 35% with increase in resources around 25% for SP and between 22% and 21% with increase in resources between 17% and 20% for DPP.

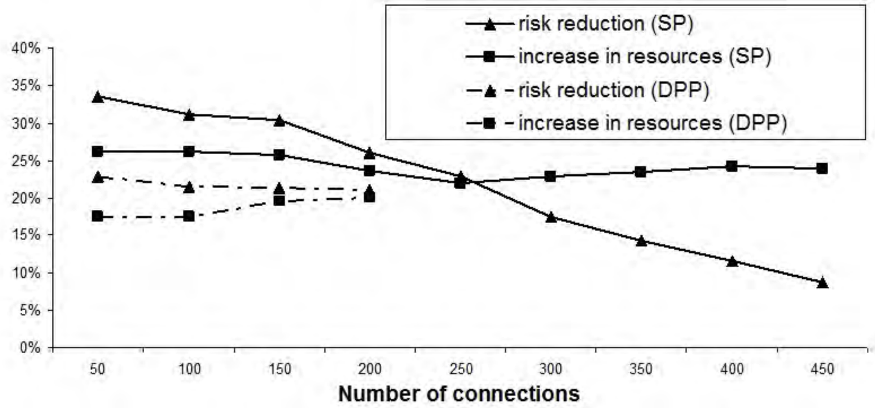


Figure 7. Risk reduction and increase in resources.

What this risk reduction really means can be shown by a sample scenario. For instance, a WMD attack on Washington DC may affect the node located there and links originating from this node. In this case, the network operator can reduce penalty by 40% by risk-aware provisioning. As another example, an attack on New York City may affect the node located there and links originated from this node. Note that, because of population-oriented traffic distribution, many connections are originated at or destined to this location. Even in this case, the network operator can obtain a 15% penalty reduction. Figure 8 shows the average penalty reduction considering all possible WMD attacks shown in Fig. 2.

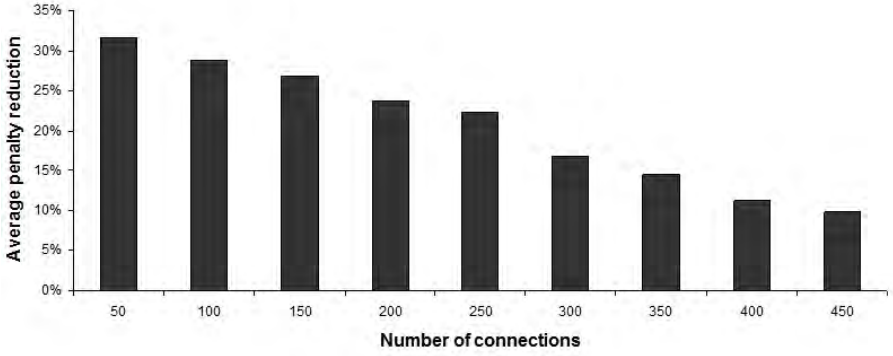


Figure 8. Average penalty reduction for single-path provisioning.

Risk-aware provisioning requires more resources than risk-unaware provisioning. However, EC can be exploited to provide better protection against WMD disruptions (e.g., avoid risky zones, whenever possible, by using the excess capacity (EC) of the links traversing safe zones), when there is a large amount of EC. When EC becomes low, traversing less-risky regions could be an acceptable risk to lower the resource consumption and to avoid capacity exhaustion. We have also studied how to exploit EC to improve network robustness [5].

A Recovery Method: WMD-Aware Reprovisioning

After an attack, some of the traffic will be disrupted. This disrupted traffic can be re-provisioned in the network by using the *excess capacity (EC)* in the undamaged areas [5]. Note that *excess capacity* is an unused capacity which any operational network has in order to accommodate traffic fluctuations and to avoid capacity exhaustion. While reprovisioning, we should also consider correlated cascading failures which might disrupt some of the connections. For instance, a WMD attack may cause power outages because of disruption on the electric grid (or technical problems in a power plant). In this case, some of the network resources will consume electricity from generators for a while. However, the diesel supplies for generators are limited because of the attack and generator-dependent network resources will suffer from power outages. Another example might be correlated sequential WMD attacks where, after a WMD attack, there may be possible secondary targets. Thus, we also need to reprovision connections under the risk of correlated cascading or sequential failures.

We can use the proposed ILP and heuristic solutions used for WMD-aware provisioning for reprovisioning by properly adjusting the inputs. The given topology, *G*(*V*,*E*), should be replaced by the damaged topology, where some of the network resources are down by the attack. The set of possible attacks *N* now denotes the set of possible correlated cascading or sequential failures. We prefer reprovisioning a subset of connections (those traversing a damaged area and those under the risk of correlated failures), denoted by T_r, instead of reprovisioning all (*T*) to keep reprovisioning time as small as possible. In this case, Eq. (5) changes as follow:

$$\sum_{t\in T} R_{ij}^{t} \le F_{ij}, \quad \forall (i,j)\in E \tag{14}$$

To evaluate our recovery approach, we give numerical examples for a WMD attack on Washington DC where number of connections is 450. After the attack, some of the network resources will not be available until full recovery. 43% of the connections traversing the node located in Washington DC can be reprovisioned since they are not destined to or originated from this node. Before reprovisioning, we should also consider the correlated cascading failures and possible sequential attacks. For instance, the damage on electric grid and some power station (e.g., Calvert Cliffs nuclear power plant in Lusby, Maryland, a nearby city to Washington DC) may cause large power outage which might affect the node in Princeton, NJ. Thus, there is a high risk of correlated cascading failure for this node and links in this shared-risk group (SRG) and 16% of connections traversing this node can be reprovisioned. We also consider that Los Angeles, CA, is under risk of a correlated secondary WMD attack which might affect the link between Palo Alto, CA, and San Diego, CA. All connections traversing this link can be reprovisioned (depending on excess capacity in other parts of the network).

In Fig. 9, the first column shows the penalty from the main attack (on Washington DC), correlated cascading failure in Princeton, NJ, and a sequential attack on Los Angeles, CA, if the network operator does not reprovision any connection. If the network operator reprovisions only the connections traversing Washington DC and does not consider reprovisioning connections under the risk of cascading failures or sequential attacks (shown in second column), the penalty caused by the attack on Washington DC is reduced by 26%. However, the total penalty reduction (considering all failures) will be 9.5%. In addition to 26% penalty reduction in Washington DC area, risk-aware reprovisioning (shown in third column) provides 18% penalty reduction for a cascading failure in Princeton, NJ, and 100% penalty reduction for the connections traversing Los Angeles, CA area. The total penalty reduction is increased to 28.7%.

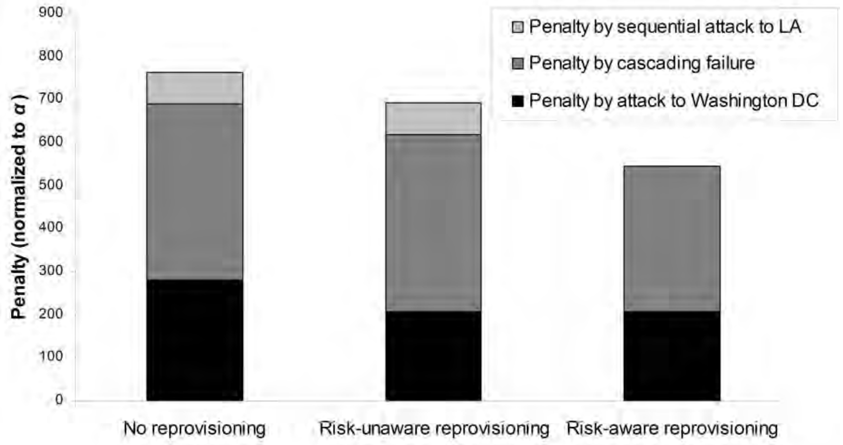


Figure 9. Penalty in case of correlated cascading failures and sequential attacks.

Network Adaptability by Exploiting Degraded-Service Tolerance

To alleviate a WMD attack's impact, new measures can be taken, depending on service requirements since the nature of the network changes dramatically as available resources decrease during disasters. We develop a metric, called degraded-service tolerance, which can reduce protection cost, reduce network disruption, and support maximal carried traffic in case of disasters. Some services are sensitive to the amount of capacity provided, while others (e.g., video streaming or file transfers) can operate with reduced bandwidth. "Degraded service" refers to the reduced amount of resource allocation for a service vs. its normal operation requirement. So, even with "degraded service", some services can still achieve lower but acceptable quality.

Degraded-service-tolerant connections can be admitted and recovered with reduced bandwidth under resource crunch caused by WMD attacks. A limitation of prior studies exploiting degraded service is that they do not adapt their resource allocation based on the network state. Our scheme [6] re-assigns resources among connections by leveraging their degraded-service tolerance. We make two major contributions besides exploiting degraded-service tolerance as partial-protection schemes do. First, our disaster-aware provisioning considers the decision process in the aftermath of disasters separately from the provisioning step, as it also considers the degraded-service tolerance of incoming connections. Second, during disasters, we allow provisioned connections to also be part of the resource allocation optimization. To offer the affected connections acceptable levels of services, unaffected connections can be degraded, re-routed, or even halted depending on the importance or profitability characteristics.

Degradation-Aware Adaptive Resource Allocation

The existing solutions perform static resource allocation, e.g., once admitted, no further actions are taken on a connection unless there is a fault that affects it. Such non-adaptive approaches result in suboptimal solutions. We propose and study the characteristics of a novel resource allocation framework (service provisioning and reprovisioing) to enable the network to be robust to WMD attacks by adaptively responding to the changes in the network state. This is a new approach to (re)distribute resources among existing connections. Our work can coexist with prior disaster-aware provisioning schemes. We aim to minimize blocking rate (unadmitted connection ratio over all requests), dropping rate (of losing connection), and disruption rate (rate of reprovisioning) of connections during disasters. Also, we aim to provide the best service possible to connections with remaining network resources by rearranging the resource allocation.

We exploit the degraded-service concept to combat disasters as follows:

(1) Accept degraded services during call admission to increase service acceptance rate, and if necessary, degrade existing connections,
(2) Degrade existing connections to reduce dropping rate, and
(3) Apply an upgrade process to restore degraded connections to full service whenever possible.

Figure 10 shows our proposal with three steps: *provisioning*, *recovery,* and *upgrade*. The *provisioning step* is applied only when a disaster occurs; else disaster-aware provisioning with full service is used. Specifically, we deal with the limitation of available resources caused by a disaster during the admission process by accepting connections with degraded service and/or adapting the network by degrading existing connections to release some resources for incoming connections. *Recovery step* kicks in when some connections get disrupted due to a disaster failure; then we try to release some capacity to avoid them to get dropped. Finally, the *upgrade step* is triggered when new resources are reactivated after the repair of some network elements, to provide the best service to connections within available network resources.

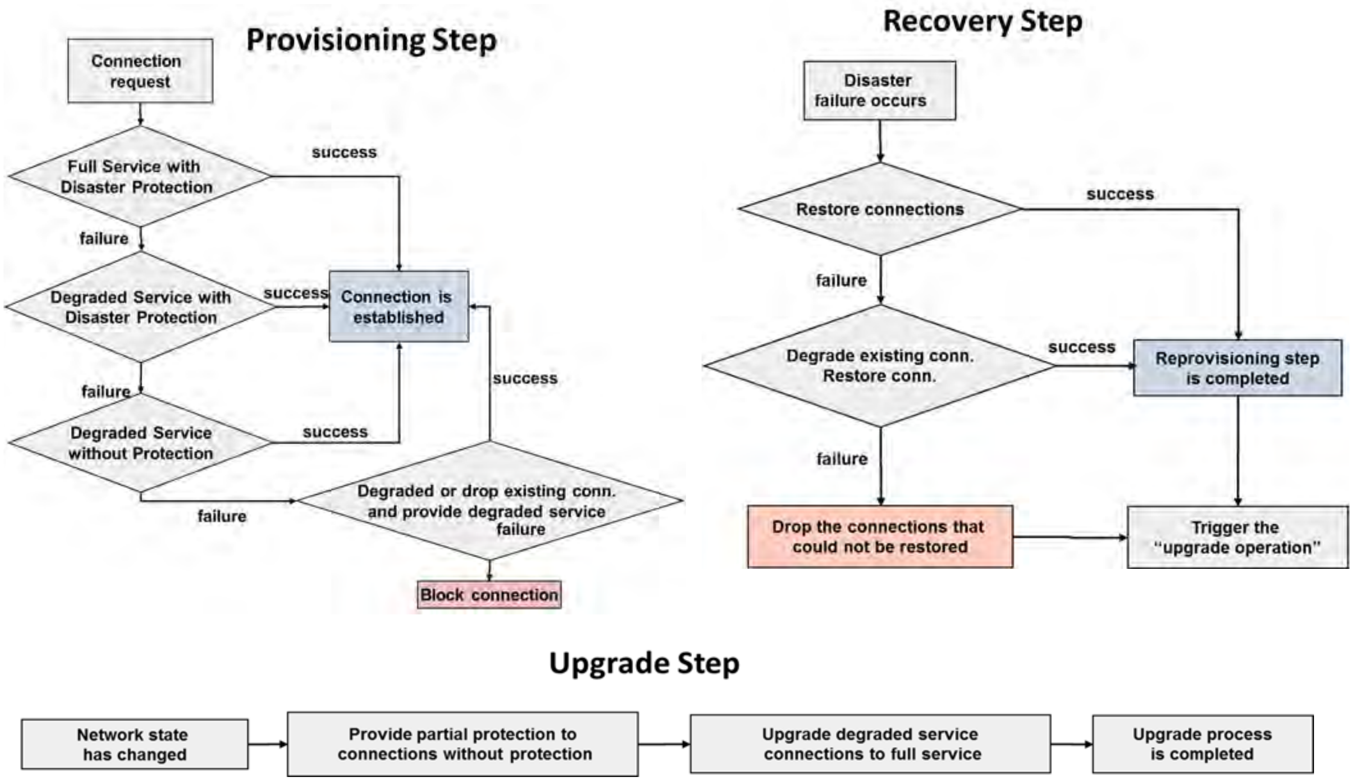

Provisioning Step
Connection request
Full Service with Disaster Protection
success
failure
Degraded Service with Disaster Protection
success
Connection is established
failure
Degraded Service without Protection
success
success
failure
Degraded or drop existing conn. and provide degraded service
failure
Block connection
Recovery Step
Disaster failure occurs
Restore connections
success
failure
Degrade existing conn. Restore conn.
success
Reprovisioning step is completed
failure
Drop the connections that could not be restored
Trigger the "upgrade operation"
Upgrade Step
Network state has changed
Provide partial protection to connections without protection
Upgrade degraded service connections to full service
Upgrade process is completed

Figure 10. Operational steps.

For illustration, consider three service levels which can be provided to a connection by multipath routing (Service levels 1-3 in descending order of quality). These three service levels can be seen in Fig. 11 for a connection request with 120B full service and 40B degraded service requirement. This connection can be admitted with any service level depending on the network state. Also, when additional resources are required during reprovisioning or admission (Stages 1 and 2), by tearing down paths, we can degrade this connection gracefully. For Stage 3, simply adding paths will enhance the service level of the connection. Without causing any disruption, by tearing down or adding some paths to a connection, we can adapt the network according to its current state. Our proposed solution is not restricted with multipath provisioning and can be applied to any existing provisioning scheme.

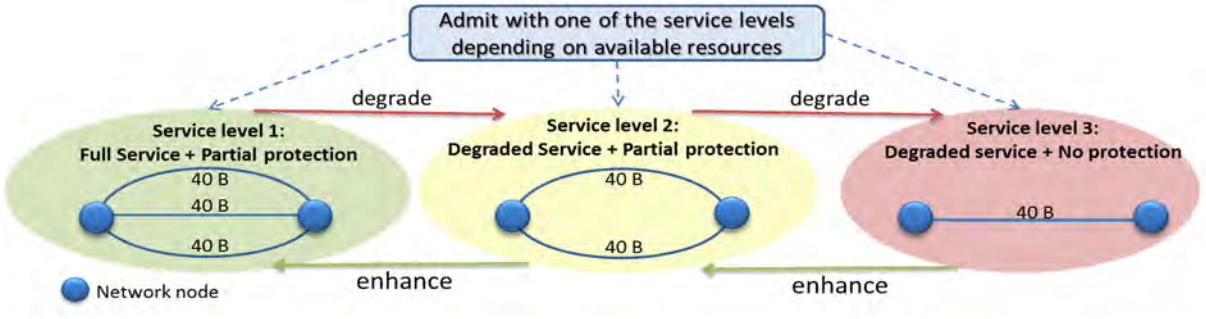

Admit with one of the service levels depending on available resources
degrade
degrade
Service level 1: Full Service + Partial protection
40 B
40 B
40 B
Service level 2: Degraded Service + Partial protection
40 B
40 B
Service level 3: Degraded service + No protection
40 B
Network node
enhance
enhance

Figure 11. Benefits of multipath provisioning by exploiting degraded-service tolerance of connections.

Case Study

We analyze the steps (admission, reprovisioning, and service upgrade) by comparing three different approaches with heuristics to gain insights about the problem. Degraded-Service (DS) and

Extreme-Degraded-Service (EDS) are the proposed approaches whereas Full-Service (FS) is the traditional approach.

Full-Service (FS) Scheme: Degraded-service is not exploited during WMD disruption, i.e., no reallocation of network resources is allowed and connections are only accepted with full service.

Degraded-Service (DS) Scheme: At admission, connections can be accepted with the best possible service level between the ranges of degraded service to full service. During recovery, we reprovision disrupted connections by gradually degrading the service level, until it gets satisfied. In this scheme, reallocation of network resources among existing connections to release resource is not allowed.

Extreme-Degraded-Service (EDS) Scheme: At both admission and reprovisioning, this scheme allows rearrangement in the network by relocating or degrading existing connections to release resources for incoming and disrupted connection after trying the steps in degraded-service scheme.

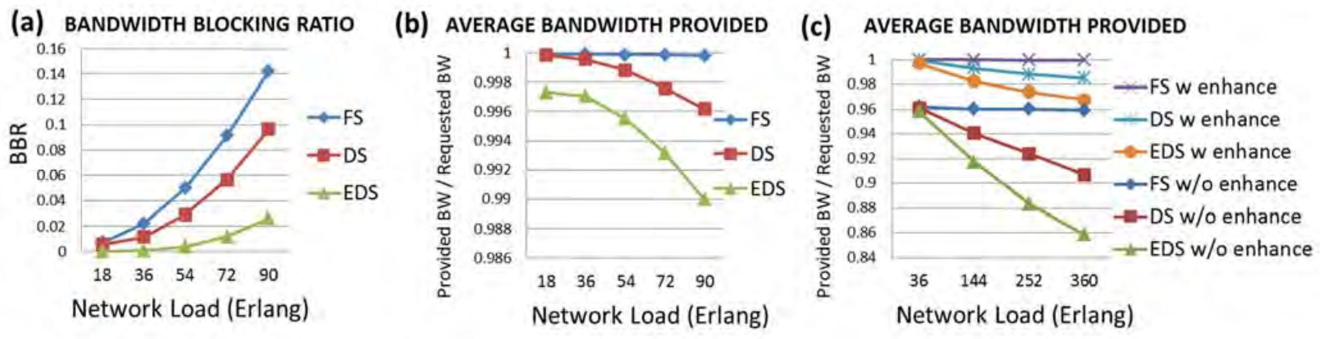


Figure 12. Comparison of the proposed schemes (DS and EDS) vs. a non-adaptive provisioning scheme (FS).

Figure 12(a) compares bandwidth-blocking ratio (BBR) (i.e., amount of rejected bandwidth over total requested bandwidth) of Full-Service (FS) with our degraded-service approaches. At low loads, there is no significant difference as acceptance ratio is high, but at high loads, our degraded-service schemes outperform the traditional scheme by decreasing BBR significantly. Since our schemes are much more flexible in terms of resource allocation, we can serve more connections with the same amount of bandwidth compared to traditional scheme.

Since the fraction of connections receiving degraded service is higher than that in FS (where connections only receive degraded service if they get disrupted due to a failure), the average bandwidth provided is slightly lower than in FS (Fig. 12(b)). Figure 12(c) shows the benefit of the service upgrade phase, and to observe it, the connection arrival rates for this figure are higher than the others as, at low load, disrupted connections can get reprovisioned with full service.

Network Adaptability under Resource Crunch: Availability and Revenue Maximization

Traffic is increasing at a very fast pace in backbone networks due to several reasons among which the primary being the popularity of cloud services, smart devices, video applications, etc. On top of that, there may be temporary bursts of traffic due to diurnal variations. Moreover, networks need to adapt to disruptions due to disasters, such as WMD attacks. These situations, where traffic demand is more than available resources, are known as resource crunch.

On the other hand, the network has to support different types of services which can be broadly categorized as real-time (e.g., video streaming, voice calls, etc.) and non-real-time (e.g., email, file transfer, etc.). They have different Service-Level Objectives (SLOs) which specify user requirements, e.g., bandwidth (BW), availability, etc. Under resource crunch, some services such as video streaming can be degraded (e.g., provided reduced bandwidth) as long as they are in the tolerable range. As an example, 2 Mbps is required for high-quality video telephony which can be degraded to 25 kbps if there is not enough resources.

Besides availability, maximizing revenue is also important from the network operator's point of view. Existing provisioning schemes treat all service types (web browsing, email, video, etc.) equally under resource crunch with no regards to the revenue generated by (or the importances of) the services.

The contribution of this work is in providing a mathematical approach, which can handle non-linear revenue models, to maximize revenue of service providers [7]. This work also provides well-defined revenue models and suitable algorithms to maximize revenue (increasing availability being the secondary objective) and to maximize availability (increasing revenue being the secondary objective) while satisfying quality of service (QoS) for different service classes for those models.

Revenue Models

Usage-based revenue model: In this model, revenue generated by a service is proportional to the usage of resource (BW in this case). Services with more flexibility in their bandwidth demand can be degraded in order to make room for new requests. This facilitates more availability and more revenue. So, this model grants discount to those services with higher flexibility. We define flexibility as:

$$F = \frac{SLO\ (max) - SLO\ (min)}{SLO\ (min)}$$

Revenue generated per unit time by a service is given by:

$$R = \frac{K}{CF+1}(\sqrt{ABW - SLO(min)} + SLO(min))$$

where *K* and *C* are constants and *ABW* =Allocated Bandwidth. Note that the revenue function is concave which implies that, when a service uses more bandwidth, it pays less per unit bandwidth. This concept of volume discount will also be inherent to the next revenue model.

Priority-based revenue model: In this model, packets sent by users are marked according to priorities and users are charged according to their chosen priority. But users have no control over resource (BW) reservation in this model. This is one of the two drawbacks of this model, the other one is that low priority services may constantly be deprived in presence of high-priority services. This model was less popular until recently when many service providers have been thinking of prioritizing their users to provide better services to the premium users under resource crunch. We modify the model to get rid of the first drawback, i.e., the services are provisioned according to their priorities but their QoS is ensured once provisioned. Let us assume priority is denoted by *P*. The suggested revenue model is as follows:

$$R = AP(\sqrt{ABW - SLO(min)} + SLO(min))$$

where *A* is a constant.

Mathematical formulation

In this work, we propose interior point method for optimally redistributing the resources among a set of services. The variables used are:

1. $f_i(R,k)$: Revenue generated per unit time for *i*th service of type *k* and resource (BW) *R*
2. C_l: Capacity of link *l*
3. $SLO_k(\mathrm{min})$: Minimum BW demand for service of type *k*
4. $SLO_k(\mathrm{max})$: Maximum BW demand for service of type *k*
5. S_l: Set of services provisioned on link *l*
6. R_i^k:Resource allocated to *i*th service (type *k*)

Objective:

$$Min \qquad -\sum_{i=1}^{N} f_i(R,k)$$

where *N* is the number of services.

Constraints:

$SLO_k(\min) \leq R_i^k \leq SLO_k(\max)\ \forall i = 1,2,\dots,N; i \neq new$ (15)

where *new* is the index of the newly arrived request.

$R_{new}^k \in [SLO_k(\min), SLO_k(\max)]\ \cup 0$ (16)

$\sum_{i \in S_l} R_i^k \leq C_l\ \ \forall l$ (17)

The objective is to maximize revenue. Equation (15) ensures that all the requests which are already provisioned are allocated BW satisfying their QoS so that none of the already-provisioned services are dropped. Equation (16) signifies that the new request can either be rejected or will be allocated BW satisfying its QoS, if accepted. Equation (17) is the capacity constraint for each link. While implementing the algorithm, instead of Eqn. (16), we use:

$0 \leq R_{new}^k \leq SLO_k(\max)$ (18)

i.e., we allow the optimizer to allocate as low as 0 unit BW to the new request. If the new request is allocated less than its corresponding $SLO_k(\min)$, then we reject the new request and redistribute the capacity (if any) allocated to it.

Though this kind of mathematical formulation produces good results, they are computationally expensive. This may be an issue for a dynamic scenario where network state is changing very frequently. So, we present heuristic approaches which, although sub-optimal, are suitable for dynamic scenarios.

Algorithm: Greedy Redistribution algorithm

The provisioning scheme for a new request when there is no resource crunch is as follows:

1. Search for *k* shortest paths
2. Calculate available capacity (*Path_cap*) on each path.

   Free_cap = Available capacity on a link without degrading any service

   cap_curr ← *Free_cap* of all the links

   Path_cap = Minimum *Free_cap* on all links of a path

3. Choose path with max. *Path_cap* (single-path routing)
4. If *Path_cap* > *SLO* (*max*), allocate *SLO* (*max*) Else, allocate *Path_cap*

When a new request arrives under resource crunch, the idea of degrading some existing services and redistributing resources comes into play. The following concepts are used in the algorithm:

Available Capacity after Degradation (*ACD*): *Free_cap* + Capacity available from degrading other services on a link to corresponding *SLO (min)*'s*.*

resalloc: Resource allocated to the services at present.

When a request arrives under resource crunch, Steps 1-2 decide the possible path for the new request. Step 3 decides whether to provision the request depending on availability of enough resource. Step 4(a) ((b)) degrades *adjacent services* (*AS*) of the new request and redistributes the resources among *AS* to maximize availability (revenue) ((revenue (availability) maximization being the secondary objective). Step 4(c) is a mathematical optimization to redistribute resources among all the services to achieve a trade-off between availability and revenue maximization. After performing Step 4(a) or 4(b) or 4(c), Step 5 checks whether any other service can be allocated more resources in decreasing order of *V*.

```
//Arrival
1. Calculate available capacity on every path after
   degradation (Path_ACD)
   Path_ACD=Minimum of ACD of all links of a path
2. Choose path with maximum Path_ACD
3.  if Path_ACD < SLO(min)
block the request
    end if
    if Path_ACD > SLO(min)
    perform either of 4(a), 4(b), or 4(c).
4(a). //This algorithm maximizes availability (revenue maximization being the secondary objective)
      allocate SLO (min) to the new request
      excess = Path_ACD - SLO(min)
      sort AS in descending order of a parameter V (defined later)
      EA_list ←calculate_EA(AS, resalloc)
      //Calculate EA for the services in AS
          for (all services in AS)
              Excess_allotted =min (EA, ACD of the
                links that the particular service shares with the path of the new arrival)
                resalloc (service) = SLO(min) +
                                        Excess_allotted
                Update ACD on all the links
                end
                Update cap_curr
 4(b). //This algorithm maximizes revenue (maximizing availability being the secondary objective)
       resalloctemp=resalloc;
       Perform Step 4a
       Compare revenue generated per unit time according to resalloctemp ($R_1$) and resalloc ($R_2$)
       if ($R_1 > R_2$)
         reject the new arrival
         resalloc=resalloctemp
       end if
4(c). //This algorithm uses mathematical optimization framework
        Perform optimization explained in section IV among
        all the services currently provisioned in the network
        if ($R^k_{new} < SLO_k(\mathrm{min})$)
           reject the new request
        end if

5.   all = {sorted array of all services in descending order
             of V}
        for all services in all
          if (resalloc(service) = SLO(max))
            continue
         else if (min (Free_cap of links on service path) = 0)
continue

else resalloc(service)=resalloc(service)
      +min(Free_cap of links on service path)
      Update cap_curr
end if
   end for
end if
// Departure
6.   Free the resources being used by that service on its
      path
7.   Update resalloc, cap_curr
8.   Execute Step 5
```

When there is service departure, Step 6 frees the resources being used by that service on its path. Step 7 updates *resalloc* and *cap_curr*. Step 8 checks whether any other service can be allocated more resources in decreasing order of *V* (similar to Step 5).

The parameter *V* depends on revenue model.

Usage-based revenue model: *V* of a service is dependent on two factors: a) Increment (*I*) in the revenue if *ABW* is increased by one unit and b) Resource consumed by it. If there are two services that have the same *I,* the one which is provisioned over more number of links will end up using more BW. So we define *V* as:

$$V = \frac{I}{\#links\ on\ which\ the\ service\ is\ provisioned}$$

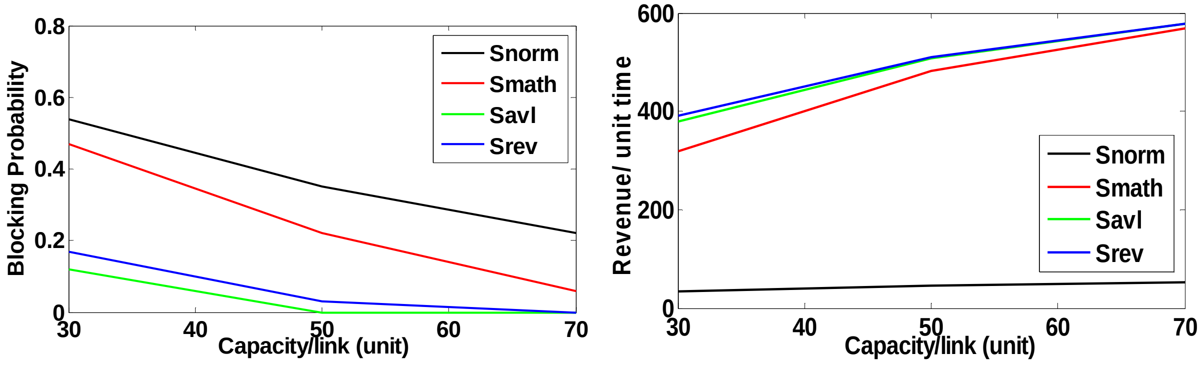
Priority-based revenue model: For a priority based revenue model, the services are allocated BW according to their priorities. So:

$$V = P$$

Note that, in the heuristic approach, we limit the degradation and redistribution only within the *adjacent services* of the new request. This approach, while being suboptimal, demands less computational power and time and hence, more practical.

Illustrative Numerical Example

i) Snorm: Scheme without degradation and redistribution (rejects a request under resource crunch),
ii) Savl: Scheme implementing Step 4(a) of the algorithm,
iii) Srev: Scheme implementing Step 4(b) of the algorithm, and
iv) Smath: Scheme implementing Step 4(c) of the algorithm.



(a)

(b)

Figure 13. Comparison of four schemes for usage-based revenue model.

Figure 13 indicates that Snorm performs worst in terms of both availability and generated avenue, as it does not exploit degradation and redistribution of resources. The heuristic approaches perform better than the mathematical approach. This is because the way the problem is formulated; the optimization incurs some penalty whenever it allocates the new request less than the corresponding *SLO* (*min*). So, the model tries to allocate the new request more than its *SLO* (*min*) unless it is really necessary to do otherwise. This is a tradeoff between the availability and revenue generation which is indicated in Fig. 13. These schemes produce similar results for priority-based revenue model also.

C. WMD survivability in cloud networks

While traditional approaches presented above focus on protecting links and nodes (routers, switches, etc.) to protect end-to-end connectivity, the shifting paradigm towards cloud computing and virtualization requires that we protect the data (or content). In today's networks, more than 90% of traffic is due to content dissemination [8]. Government, military, and commercial networks have also adopted these new trends, e.g., Department of Homeland Security (DHS) is building its Private Cloud to manage sensitive information [9], to provide better service at a lower cost, and to expand capacity at a moment's notice. Thus, data/content protection/restoration against WMD attacks is very important, along with its reachability from the end users.

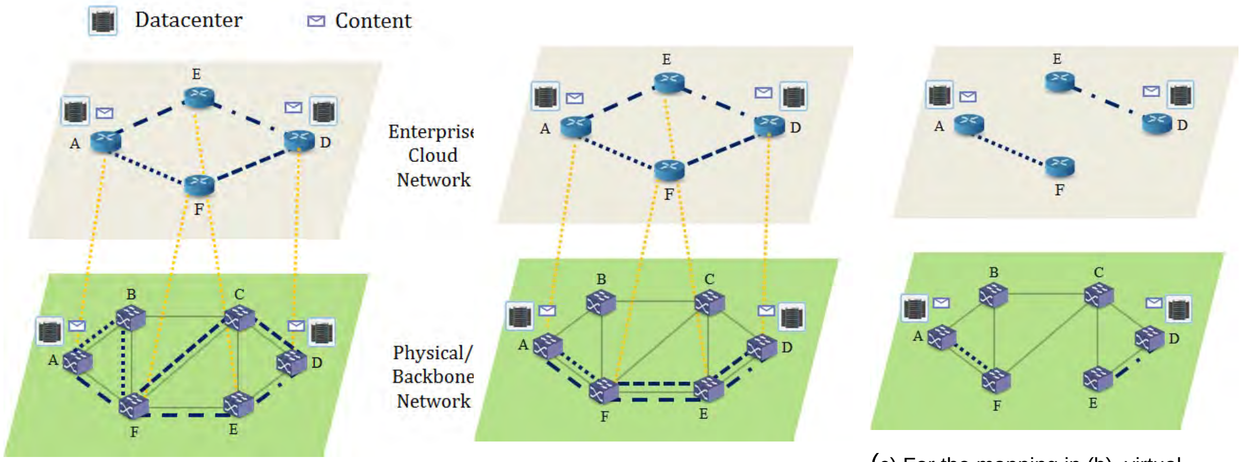
Disaster-Resilient Cloud: From Network Connectivity to Content Connectivity

The majority of today's Internet traffic is generated by cloud services (e.g., data-in-cloud, video streaming, etc. [8]) which require users to access and transfer contents from datacenters. As cloud services become an integral part of our daily life, disruption of content/service availability – especially in case of large-scale WMD attacks – can cause major upheaval in the economic well-being and security of our nation. Telecom networks and the Internet were originally designed to provide end-to-end communications which can survive failures. Thus, network connectivity (i.e., reachability of any network node from all other nodes) has been the traditional survivability need against failures. However, due to the shifting paradigm towards cloud services, we now need to ensure that, at any moment in time and under any failure condition, users still have access to (e.g., a copy of) their data/content, even if the network is disconnected. In other words, we have to ensure reachability of a content from any point of a network, which we call Content Connectivity (in contrast to network connectivity), after a failure [10]. We present the new protection paradigm of content connectivity and propose a mathematical model to compute virtual network mapping and content locations to ensure content connectivity against large-scale WMD/disaster failures. We also developed novel algorithms for virtual network mapping and content placement to ensure content connectivity in the cloud.

A cloud network is generally an overlay (i.e., virtual/logical) network mapped over a physical network, e.g., an enterprise IP network over a carrier's fiber-optic network. Virtual network (VN) mapping is defined as the assignment of VN resources to the elements of the physical network, e.g., virtual links are created using optical layer lightpaths. Typically, multiple VNs (each belonging to a different military entity or group) may be overlaid on a common physical network (e.g., a backbone network). Providing failure-resilient content connectivity depends on the selection of an appropriate VN mapping, because even a single failure in the physical network may cause multiple failures in an overlay network which might make it disconnected.

Figure 14 explains the concept of VN mapping for a two-layer network, where a four-node cloud network with two datacenters (to simplify we assume that both host the same contents) is mapped over a six-node optical (physical) backbone network. In this small example, for illustration purposes, we consider that a random attack can damage a single link at a time.

Figure 14(a) shows a VN mapping that ensures network connectivity after a failure. The virtual links A-E, A-F, D-E, and D-F are mapped on physical-layer paths A-F-E, A-B-F, D-E, and D-C-F, respectively. In this simple mapping example, none of the physical links carries more than one virtual link. Thus, in the event of a physical link failure, only one virtual link might be damaged, and it is guaranteed that the virtual (cloud) network will remain connected.



(a) Survivable mapping ensuring network connectivity. 16.67% more resource usage than the mapping in (b) which ensures content connectivity.

(b) Survivable mapping ensuring content connectivity, but does not ensure network connectivity.

(c) For the mapping in (b), virtual network gets disconnected if physical link F-E is down, but user from any node can reach the content.

Figure 14. Survivable cloud network mapping.

The mapping shown in Fig. 14(b) ensures content connectivity, i.e., even if the network is disconnected after a failure, each segment of the disconnected network holds a copy of the content. For example, if physical link F-E is damaged, virtual links A-E and F-D will also be damaged as the optical paths mapping these two virtual links go through physical link F-E. Thus, the network will be disconnected into two segments {A, F} and {D, E} as shown in Fig. 14(c). Nonetheless, as both datacenters at A and D have replicas of the content, any user from any node of the network can still access the content. Similarly, if physical link A-F is damaged, virtual links A-E and A-F will be damaged, causing two disconnected segments {A} and {D, E, F}, but content connectivity is maintained. Note that the mapping ensuring network connectivity (Fig. 14(a)) uses 16.67% more resources (computed as the total number of wavelength-links used) than the mapping ensuring content connectivity (Fig. 14(b)).

Now consider that we have the same cloud topology as in Fig. 14, but the physical topology does not have link C-F. In this scenario, it turns out that there is no possible mapping that ensures network connectivity after any failure, though content connectivity can still be ensured with the same mapping as in Fig. 14(b). This example on a small problem instance shows the potential of content connectivity, namely:

a) Maintaining network connectivity may not always be possible after failures, but content connectivity can help to provide continued services in such scenarios, and

b) Ensuring content connectivity may require less network resources than network connectivity.

Schemes to Ensure Content Connectivity

We consider that a military group's (overlay/virtual) network should be mapped on a backbone network by ensuring content connectivity after a possible failure with minimal resource usage. Maintaining content connectivity also requires efficient content placement by exploiting the a-priori knowledge of network mapping and potential locations vulnerable to WMD failures in the network. Note that different failure events may create different sets of disconnected network segments. Thus, the problem of ensuring content connectivity is non-trivial.

As stated before, a cloud-network link is created (mapped) using a physical-layer path. We formulate the problem of ensuring content connectivity in a cloud as follows:

Given:

1. Physical (optical network) topology
2. Cloud (virtual) topology
3. Set of datacenters
4. Datacenters' storage capacity
5. Set of contents
6. Set of contents required by each user (node)
7. Set of WMD zones

Output:

1. Network mapping (i.e., physical layer path for each virtual link)
2. Replica locations of each content

Objective:

Minimize resource usage

Constraints:

1. A physical-layer path (i.e., lightpath) conserves its flow along the path
2. From a node, it is possible to access a required content from at least one datacenter after any WMD failure
3. A virtual link will be damaged if it traverses a physical link damaged by a WMD attack
4. Storage constraint for datacenters
5. Capacity constraint for network links

We modeled the optimal mapping problem as an Integer Linear Program (ILP) [10]. But the ILP, though it provides an optimal solution, is time-consuming and does not scale well due to its computational intractability. For large networks, we developed algorithms to solve the problem in two stages: 1) cloud network mapping and 2) content placement on the cloud. Details follow.

Cloud Network Mapping

We map virtual links one at a time. At each step, we choose the virtual link whose mapping will increase most the average connectivity from a node to the datacenters (DCs). We define two metrics for this purpose – Average Datacenter Connectivity (ADC) and Maximum Disconnection from Datacenters (MDD) – as follows:

$$ADC = \text{Avg. \# DCs reachable from a node after a failure}$$

$$= \frac{\sum_{\mathrm{n \in N, z \in Z}} \text{\# DCs reachabe from node n after WMD attack z occurs}}{\mathrm{n \times z}}$$

where *N* is the set of nodes and *Z* is the set of WMD attack zones, and

$$\mathrm{MDD} = \max\{\mathrm{d_n : d_n} = \max \text{\# DCs unreachable from node n after a WMD failure}\}$$

While ADC tries to increase the average connectivity to datacenters (i.e., number of datacenters reachable from a node) after a failure, MDD's goal is to reduce the maximum number of datacenters disconnected from a node after a failure. At each step, we choose the virtual link for mapping which increases ADC or decreases MDD most. We compute two possible mappings – one using ADC and another using MDD – and choose the result that provides more connectivity with fewer resources. The procedure is shown in Algorithm 2.

Content Placement

Algorithm 2 provides us with the mapping of the cloud network over the backbone network. Once we have the mapping, we can determine how the cloud network might get disconnected due to a failure. For each potential failure scenario, there is a set of resulting connected network segments. Considering all failure scenarios, we can build a set S of all possible connected segments. Note that some of the segments will overlap as they are created from different failure scenarios.

Also, some segments may not require a specific content (i.e., none of the nodes in the segment requires that content). Thus, while placing a content, we consider only those segments where the content is required. Note that a datacenter node can be a member of multiple segments. While placing contents, we need to ensure that each segment has at least one replica of each of the contents it requires. Basically, from the set of datacenter nodes, we have to select minimum number of nodes to place a content such that every connected segment that requires the content has a replica of the content. This way, we ensure content connectivity while minimizing the number of replicas per content. This problem is known as the “Hitting Set” problem (equivalent to Set Cover problem) which is NP hard. Algorithm 3 presents a heuristic to find replica locations of a content while minimizing number of replicas and load-balancing among datacenters. We take one content at a time and find its replica locations ensuring content connectivity. The order of the contents for replication can be based on their importance, i.e., some contents are mission-critical or more revenue generating than others.

Algorithm 2: Cloud Network Mapping

Input: Cloud (virtual) network, backbone/ physical network, *k*-shortest paths through physical network for each virtual link.

Output: Mapped virtual links over physical network.

```
repeat
    If there are ‘$n$’ unmapped virtual links and ‘$k$’ shortest paths for each virtual link, then there
    are $n \times k$ options to choose a path. Select a path (i.e., the mapping of a virtual link) that
    increases ADC most.
until
    All virtual links are mapped.
If the mapped network does not ensure connectivity to at least one datacenter after a failure,
change mapping of a link that increases ADC most. Repeat this step maximum ‘$n$’ times or until
required connectivity is ensured, where ‘$n$’ is the number of virtual links.
Store the mapping into MAP1.

repeat
    If there are ‘$n$’ unmapped virtual links and ‘$k$’ shortest paths for each virtual link, then there
    are $n \times k$ options to choose a path. Select a path (i.e., the mapping of a virtual link) that
    decreases MDD most.
until
    All virtual links are mapped.
If the mapped network does not ensure connectivity to at least one datacenter after a failure,
change mapping of a link that decreases MDD most. Repeat this step maximum ‘$n$’ times or until
required connectivity is ensured, where ‘$n$’ is the number of virtual links.
Store the mapping into MAP2.

if ADC(MAP1) > ADC(MAP2)
    Return MAP1.
else if ADC(MAP1) < ADC(MAP2)
    Return MAP2.
else if cost(MAP1) < cost(MAP2)
    Return MAP1.
else
    Return MAP2.
```

Algorithm 3: Placement for Content *c*

Input:

S: Set of network segments that require *c*.

D: Set of datacenter nodes.

W_d: Available storage capacity of datacenter $d \subset D$.

W_c: Storage requirement for content *c.*

Output: $R \subset D$: Set of replica locations for *c.*

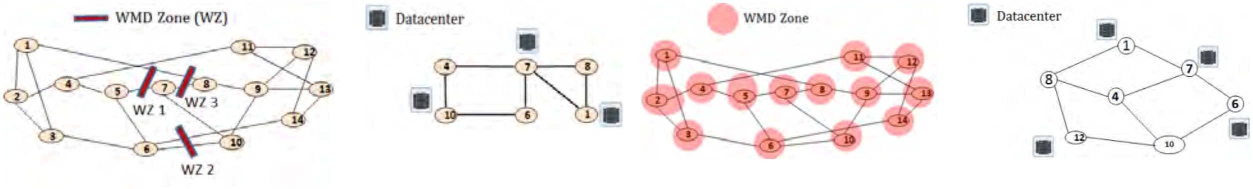
```
1. $R = \phi$.
2. while $S$ is not empty do
3.     Choose $d \in D$ minimizing $\frac{1}{W_d \times \text{number of segments covered by } d}$.
       $R = R \cup \{d\}$, $W_d = W_d - W_c$.
4.     Remove $d$ from $D$ and $s$ from $S$ such that $d \in s$.
5. end while
```

Illustrative Examples

We present illustrative results by solving the ILP model and the heuristic algorithms on two sample network topologies shown in Fig. 15, where virtual topology 1 (2) will be mapped over physical topology 1 (2). For ease of illustration, we assume that all contents have the same storage requirement and same importance. We first present results from sample topology 1 (physical and virtual) using ILP. We consider two contents, maximum two replicas per content, and four WMD attack scenarios – Scenario 1: all single-link failures; Scenario 2: all single-link failures and WZ1; Scenario 3: all single-link failures, WZ1, and WZ2; and Scenario 4: all single-link failures, WZ1, WZ2, and WZ3 (Fig. 15). In Fig. 16(a), we report the mapping cost calculated as total number of wavelength-links used for mapping. We see that, in the first three scenarios, network resource usages to maintain content connectivity are the same, and it goes up for the fourth scenario. For network connectivity, wavelength usage in Scenario 2 is larger than in Scenario 1, while for Scenarios 3 and 4, no survivable mapping exists that maintains network connectivity. Thus, network connectivity requires more resources than content connectivity (as expected, resource usage for both content connectivity and network connectivity increase when a larger set of disaster zones are considered, but the capacity requirement for content connectivity grows less steeply than network connectivity). This result indicates that the higher the number of disasters, the higher the benefit of providing content connectivity rather than network connectivity.

Figure 16(b) shows the effect of number of replicas on network resource usage. Here, we use topology 1 (physical and virtual), but consider that every node in the virtual network holds a datacenter. We see that, with a small increase in the number of replicas, wavelength usage drops significantly. But, eventually, the wavelength usage converges to a certain value.

Figures 16(c) and 16(d) compare the performance of the exact ILP model with the proposed algorithms. For topology 1, we consider disaster Scenario 3 and two contents. We also consider topology 2 with fourteen disaster zones (Fig. 15(c)) and five contents. For ILP, we consider maximum two replicas per content for topology 1, and maximum three replicas per content for topology 2. For heuristics, we do not constraint the number of replicas per content. Figure 16(c) shows the performance of the two algorithms in terms of network resource usage. We see that the performance of the algorithms is quite close to ILP. We have done simulations on other topologies and found similar results. Figure 16(d) compares the storage resource usage of the ILP with the heuristic algorithm. As all contents have the same storage requirement, we show the resource usage as the average number of contents replicated per datacenter. We can see the similar outcome in both cases.



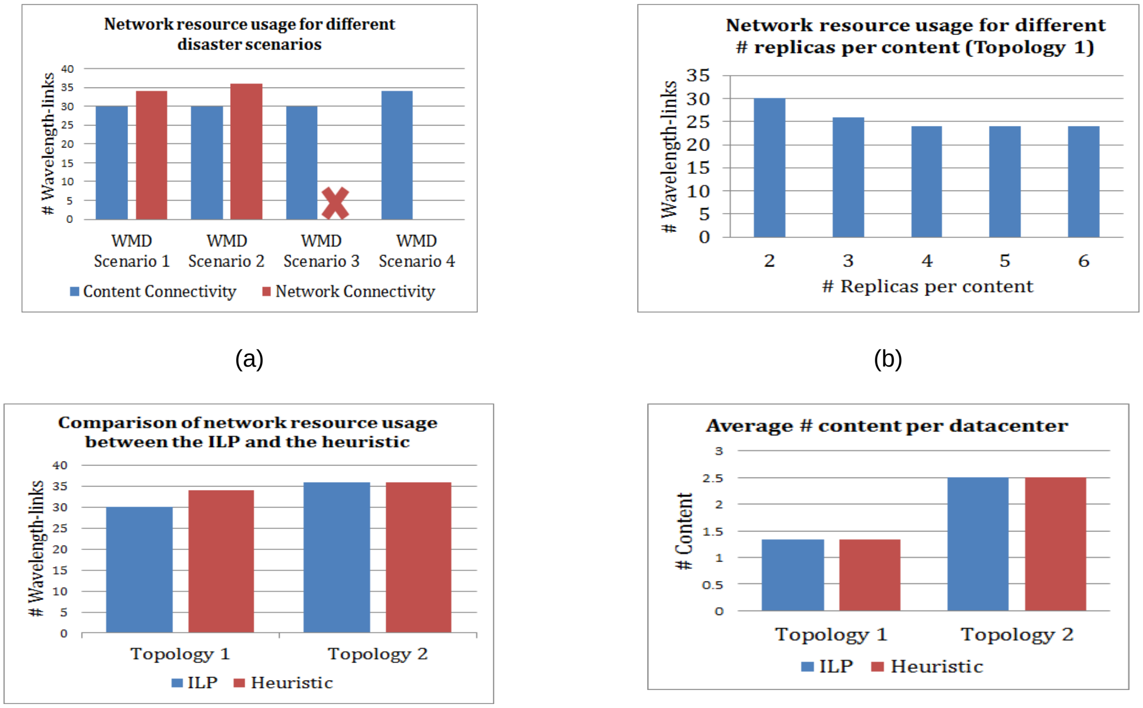
(a) Physical Topology 1

(b) Cloud (Virtual) Topology 1

(c) Physical Topology 2

(d) Cloud (Virtual) Topology 2

Figure 15. Network topologies used in the simulation studies.



(a)

(b)

(c)

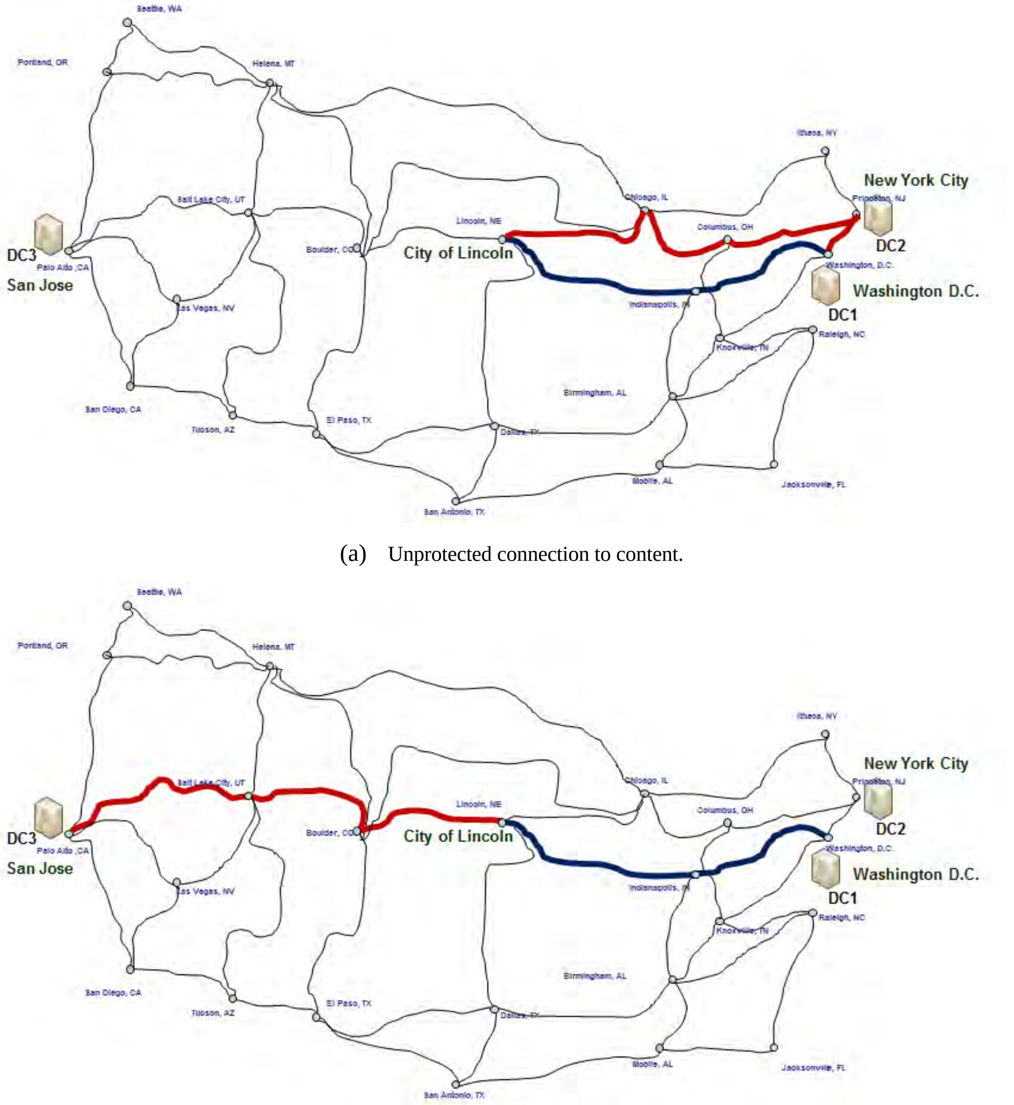
(d)

Figure 16. Simulation results.

We showed that ensuring content connectivity requires both efficient content placement exploiting the a-priori knowledge of potential vulnerable locations in the network and disaster-aware cloud network mapping over backbone/physical network. Our study shows that content connectivity is a more scalable and resource-efficient solution to ensure content/service availability after disaster failures [10].

Survivable Datacenter-Network Design and Anycasting

Typically, protection of a connection against single-link failures is ensured by providing a backup path to the same destination (i.e., datacenter), which is link-disjoint to the primary path. This scheme has been refined by the introduction of a backup datacenter, thereby adding protection against failures of a single datacenter. However, this protection scheme fails to protect against WMD attacks affecting an area that contains both primary and backup resources (either network links or datacenters). Also, the failure of a single datacenter should not cause the disappearance of a specific content/service from the whole network. Thus, three problems, namely content/service placement, routing, and protection of paths and content/service, should be addressed simultaneously while providing protection to connections in cloud.



(a) Unprotected connection to content.

(b) Protected connection to content.

Figure 17. Protection in cloud. Primary and backup paths are shown in blue and red, respectively.

Figure 17 shows an example of protection of connection and content in a cloud using primary and backup paths, which are shown in blue and red lines, respectively. There are three datacenters DC1, DC2, and DC3 located at Washington DC, New York City, and San Jose, respectively. A user from Lincoln City requests a content available at DC1. Using shortest-path routing, a possible provisioning of primary and backup paths is shown in Fig. 17(a). But both of the paths are disrupted when a WMD attack occurs in Washington DC. The content can be replicated at DC2 and the backup path can be connected to DC2, but the service might fail due to a massive WMD attack affecting both of them (by direct shock or after effects such as power outage). Moreover, this attack might damage the datacenter(s) and the content might be lost if there is no other replica of the content in the network unaffected by the attack. Figure 17(b) shows a possible provisioning to provide protection for both connection and content, where 1) the content is replicated at both DC1 and DC3, and 2) the primary and backup paths are connected to DC1 and DC3, respectively.

We developed methods to design datacenter networks while providing protection against WMD disruptions by solving the three problems (content/service placement, routing, and protection of paths and content/service) simultaneously. We consider that a lightpath is required for each request, though our work can be easily extended for arbitrary capacity requirements. We first formulated the problem into an ILP as follows:

Given:

- $G(V, E)$: Physical topology; V is the set of nodes and E is the set of directed links
- $V' \subset V$: Set of datacenter locations
- N: Set of WMD zones
- C: Set of contents
- T: Set of requests (s, c), where s is the requesting node and c is the content: $s \in V$, $c \in C$
- W: Link capacity
- K: Maximum number of replicas per content

Binary variables:

- $P_{(i,j)}^{(s,c)}$: Link (i, j) is used in the primary path for request (s, c)
- $B_{(i,j)}^{(s,c)}$: Link (i, j) is used in the backup path for request (s, c)
- $A_d^{(s,c)}$: $d \in V'$ is used as primary datacenter for request (s, c)
- $\overline{A}_d^{(s,c)}$: $d \in V'$ is used as backup datacenter for request (s, c)
- $R^{(c,d)}$: Content c is replicated at $d \in V'$
- $\alpha_n^{(s,c)}$: Primary path for request (s, c) goes through WMD zone $n \in N$
- $\beta_n^{(s,c)}$: Backup path for request (s, c) goes through WMD zone $n \in N$
- $\beta_{(i,j),n}^{(s,c)}$: Link (i, j) is used in the backup path for request (s, c) if the primary path is down due to a WMD attack occurring in $n \in N$

Integer variables:

- $\pi_{(i,j)}$: Total number of wavelengths in link (i, j) used for backup paths

Objective:

$$\min\left(\sum_{(i,j)\in E}\pi_{(i,j)} + \sum_{(i,j)\in E}\sum_{(s,c)\in T}P_{(i.j)}^{(s,c)}\right)$$

The first term above minimizes the shared backup resources and the second term minimizes the primary resources.

Flow-conservation constraints:

$$\sum_{(i,j)\in E}P_{(i,j)}^{(s,c)} - \sum_{(j,i)\in E}P_{(j,i)}^{(s,c)} = \begin{cases} 1 & \text{if } i = s \\ -A_i^{(s,c)} & \text{if } i = V' \\ 0 & \text{otherwise} \end{cases} \quad \forall (s,c)\in T, i \in V \qquad (19)$$

$$\sum_{(i,j)\in E} B_{(i,j)}^{(s,c)} - \sum_{(j,i)\in E} B_{(j,i)}^{(s,c)} = \begin{cases} 1 & \text{if } i = s \\ -\overline{A}_i^{(s,c)} & \text{if } i = V' \\ 0 & \text{otherwise} \end{cases} \quad \forall (s,c) \in T, i \in V \qquad (20)$$

$$\sum_{d \in V'} A_d^{(s,c)} = 1 \quad \forall (s,c) \in T \qquad (21)$$

$$\sum_{d \in V'} \overline{A}_d^{(s,c)} = 1 \quad \forall (s,c) \in T \qquad (22)$$

Equations (19) and (20) enforce flow conservation on primary and backup paths, respectively. The constraints ensure that, for each connection, the outgoing traffic and the incoming traffic are equal, except for the requesting node and the destination (datacenter) node. Equations (21) and (22) ensure that only one primary and one backup datacenter are assigned per connection.

Datacenter assignment and content placement:

$$R^{(c,d)} \geq A_d^{(s,c)} + \overline{A}_d^{(s,c)} \quad \forall c \in C, d \in V', (s,c) \in T \qquad (23)$$

$$\sum_{d \in V'} R^{(c,d)} \leq K \quad \forall c \in C \qquad (24)$$

Equation (23) ensures that a datacenter *d* can be assigned as primary or backup datacenter for a request (*s, c*) if and only if content *c* is replicated in datacenter *d*. It also ensures that the same datacenter is not used as both primary and backup datacenter of a request. Equation (24) constrains the number of replicas per content.

Capacity constraint:

$$\sum_{(i,j)\in E} P_{(i,j)}^{(s,c)} + \pi_{(i,j)} \leq W \quad \forall (i,j) \in E \qquad (25)$$

WMD-zone-disjoint path constraints:

$$\frac{1}{M} \sum_{(i,j)\in n} P_{(i,j)}^{(s,c)} \leq \alpha_n^{(s,c)} \leq \sum_{(i,j)\in n} P_{(i,j)}^{(s,c)} \quad \forall (s,c) \in T, n \in N \qquad (26)$$

$$\frac{1}{M} \sum_{(i,j)\in n} B_{(i,j)}^{(s,c)} \leq \beta_n^{(s,c)} \leq \sum_{(i,j)\in n} B_{(i,j)}^{(s,c)} \quad \forall (s,c) \in T, n \in N \qquad (27)$$

$$\alpha_n^{(s,c)} + \beta_n^{(s,c)} \leq 1 \quad \forall (s,c) \in T, n \in N \qquad (28)$$

Equations (26) and (27) set the value of $\alpha_n^{(s,c)}$ and $\beta_n^{(s,c)}$, respectively. Here, *M* is a large integer constant. Equation (28) ensures the WMD-zone-disjoint property of primary and backup paths.

Shared protection constraints:

$$\pi_{(i,j)} \geq \sum_{(s,c)\in T} B_{(i,j),n}^{(s,c)} \quad \forall n \in N, (i,j) \in E \tag{29}$$

$$B_{(i,j),n}^{(s,c)} \leq \alpha_n^{(s,c)} \quad \forall n \in N, (i,j) \in E, \forall (s,c) \in T \tag{30}$$

$$B_{(i,j),n}^{(s,c)} \leq B_{(i,j)}^{(s,c)} \quad \forall n \in N, (i,j) \in E, \forall (s,c) \in T \tag{31}$$

$$B_{(i,j),n}^{(s,c)} \geq \alpha_n^{(s,c)} + B_{(i,j)}^{(s,c)} - 1 \quad \forall n \in N, (i,j) \in E, \forall (s,c) \in T \tag{32}$$

Equations (29) - (32) bound the number of wavelengths used in a link for shared protection. From the objective and Eqn. (29), we find that $\pi_{(i,j)} = \max_{n\in N} \sum_{(s,c)\in T} B_{(i,j),n}^{(s,c)}$.

In this ILP, we do not explicitly include constraints to ensure protection for contents against WMD attacks. Assuming that each content is requested by at least one request, and also noting that primary and backup paths for a request are not affected by same WMD attack and connect to two different datacenters, it is ensured that content is replicated at WMD-zone-disjoint datacenters.

Relaxation of the integrated ILP

The integrated ILP discussed above is useful to analyze the characteristics of a datacenter network. However, this ILP does not scale well due to its intractability. Also, the integrated ILP solves three problems simultaneously which makes it very complex. To solve the problem for large networks, we propose two relaxations:

A. Two-Step ILP: We introduce separate but interlaced ILP formulations for content placement and routing.
B. LP Relaxations: The relaxed two-step ILP, obtained by relaxing the integrality constraint of the variables, gives us a lower bound on the optimal. Since the LP-relaxed solution may not be a feasible one, we propose heuristics to find a feasible solution from the LP-relaxed solution.

A. Two-Step ILP

Here, we consider content placement with content protection and routing with path protection as separate processes, and propose two separate ILPs for these two problems. This ILP allows us to handle more connections as it has lower complexity than the integrated one.

1) *Content Placement With Content Protection*: As we perform routing separately, we use a heuristic objective for this formulation. The objective is to place each content in datacenter locations such that the average distance from the requesting nodes to the datacenter nodes for all the requests is minimized. This objective tends to place a content in those datacenters closer to its popular region, which can reduce resource usage by primary and backup paths while routing connection requests. We pre-compute $H_{s,d}$ which is the minimum number of hops to reach node *d* from node *s*. Our ILP to solve the content placement problem follows:

Objective:

$$\min\left(\sum_{d\in V'} \sum_{(s,c)\in T} H_{s,d} R^{c,d} \right)$$

Constraints:

$$\sum_{d \in V'} R^{c,d} = K \quad \forall c \in C \qquad (33)$$

$$\sum_{d \in n} R^{c,d} \leq 1 \quad \forall c \in C, \forall n \in N \qquad (34)$$

Equation (33) bounds the number of replicas to *K*. Equation (34) ensures that no WMD-zone can have more than one replica (which can happen if at least two datacenters fall in the same WMD-zone and they both have the replica) of a content.

2) Routing with Path Protection: We derive the ILP for routing by eliminating Eq. (34) from the integrated formulation and using constant values for R_d^cs as found from Step 1. Although we eliminate only $|C| \times |V'|$ variables and $|C|$ constraints, it reduces the complexity significantly [11].

B. LP Relaxation

One way to make an ILP more scalable is LP relaxation: remove the integrality constraints on variables and then solve the corresponding Linear Program (LP). The search space of the LP includes the search space of the ILP, so LP relaxation provides a lower bound on the optimal solution. As some constraints are relaxed in the LP formulation, we may obtain an infeasible solution, particularly since this solution may give fractional values for the variables (some of which should be integers in reality). Heuristics can be used to obtain a feasible solution for the original ILP problem from the infeasible relaxed solution.

As there is no constraint on the storage capacity of data centers in the ILP to solve the content placement problem (i.e., Step 1 of two-step ILP), the placement of a content item does not depend on the placement of other contents. Thus, each item can be replicated separately. This ILP uses a heuristic objective which tends to place an item in data centers closer to its popular region. Thus, instead of LP relaxation, we propose Algorithm 4, which solves the content placement problem (one item at a time) to achieve the same objective in polynomial time.

Basically, for a specific content item and a specific data center location, we consider the requests for the item and compute the average shortest distance of that data center from the requesting nodes. We sort the data centers in non-decreasing order of the computed distances and replicate content into data centers in succession following the order of the sorted list and maintaining constraints (33) and (34).

We apply LP relaxation on the routing problem (i.e., Step 2 of the two-step ILP). We relax all the integer variables. Note that, as the relaxed solution may give fractional values for variables, we may have multiple primary paths and multiple backup paths for each connection request.

We propose Algorithm 5 to derive a feasible solution for a request (s, c) from the infeasible relaxed one. Here, *SP* is the pre-computed set of *k*-shortest paths from node s to each of the data center nodes. Using any polynomial-time search algorithm, we can find the set of primary paths, P, and backup paths, B, from the relaxed solution. Here, $\alpha_i^p (\beta_i^b)$ is equal to 1 if primary path p (backup path b) goes through WMD-zone *i*. Then, we compute $max_i(\alpha_i^p \times \beta_i^b)$ for each possible pair (p, b), $p \in P, b \in B$. If $max_i(\alpha_i^p \times \beta_i^b) = 0$ for a pair (p, b), then *p* and *b* are WMD-zone-disjoint paths. We compute set *S* which holds those pairs (p, b) having $max_i(\alpha_i^p \times \beta_i^b) = 0$. If S is not empty, we take the pair that consumes the least amount of resources (sum of primary and backup wavelengths to provide shared protection). If S is empty, we do not have any WMD-zone-disjoint pair of paths from the sets P and B. We then take the least-cost primary path from P, and compute *candidateB* as the set of paths $b' \in SP$ such that b' is WMD-zone-disjoint to *p*. If *candidateB* is not empty, we take a path *b* from it such that (p, b) pair consumes lowest cost. If *candidateB* is empty, we delete the links on *p* from the topology, and find a shortest path from the modified topology as backup path. If a backup path is not found, we delete the primary path from *P*, take the least-cost primary path from *P*, compute *candidateB* for the new primary path, and repeat the steps. If none of these works, we use Algorithm 6 to compute the paths, which we explain below.

Heuristic

Here, we explore non-mathematical heuristic approaches to solve the problem for large problem instances. We consider a static-traffic case, where all requests are known beforehand, yet, these heuristics can also be applied for dynamic traffic, where requests arrive and are processed one-by-one. For content placement with WMD protection, we propose Algorithm 4, as previously discussed. Algorithm 5 shows the heuristic to compute WMD-zone-disjoint primary and backup paths for a given request (s, c). The *k* shortest paths from node s to all data center nodes are pre-computed. Thus, if we have r replica locations for content c, we have $k \times r$ paths to be considered for request (s, c). The heuristic considers all possible pairs of paths and selects the lowest-cost WMD-zone-disjoint pair as solution. Cost of a pair of paths is the sum of the number of wavelengths used for the primary path and additional wavelengths used for shared backup path. For a comparison of the running times of the proposed methods, we refer to [11].

Illustrative numerical examples

We present illustrative results by solving the ILP formulations and heuristics on NSFNet network shown in Fig. 15(a). We specify 14 WMD zones centered at each of the nodes of NSFNet. The maximum number of wavelengths per link is 32. We compare three protection schemes: dedicated single-link failure (SLF) protection (i.e., dedicated path protection against a single link failure), shared single-link failure (SLF) protection (i.e., shared path protection against a single link failure), and the proposed shared WMD-zone failure (WF) protection (i.e., shared path protection against a single WMD-zone failure). In all three schemes, the primary and backup datacenters for a request are always different. The formulations for dedicated and shared SLF protection can be easily derived from our WMD protection model with minor changes.

Algorithm 4: Content Placement

Input:

T: Set of user requests (*s*, *c*) for content *c*

V′: Set of data center locations

h_{sd}: Minimum number of hops to reach node *d* from node *s*

Output:

L_c: Set of replica locations for content *c*

```
for each $d \in V'$ do
    $cost_d = \sum_{(s,c)\in R} h_{sd}$
end for
Sort all data centers in non-decreasing order of $cost_d$
$i = 1$, $n$ =number of data centers
while $L_c$ does not have k members and $i \le n$ do
    Add $i$'th data center from the sorted list into $L_c$ if it does not violate constraint (23)
    $i = i + 1$
end while
```

Algorithm 6: Compute primary and backup paths for a request

Input:

(*s*, *c*): User request with *s* as requesting node and *c* as content

SP: Set of *k* shortest paths from node *s* to all data centers

PP: Set of already-provisioned primary paths

PB: Set of already-provisioned backup paths

Output:

bestPrimary: Primary path

bestBackup: Backup path

minCost: Cost of the pair (*bestPrimary*, *bestBackup*)

```
$minCost = \infty$
$bestPrimary = NULL$
$bestBackup = NULL$
for each path $p_1 \in SP$ do
    if all links on $p_1$ have enough capacity, then
        for each path $p_2 \in SP$ do
            if all links on $p_2$ have enough capacity and $p_1$ and $p_2$ are WMD-
```

Algorithm 5: Routing: Compute primary and backup paths for the given request from relaxed LP solution

Input:

(*s*, *c*): User request with *s* as requesting node and *c* as content

P: Set of primary paths found for (*s*, *c*) from relaxed ILP

B: Set of backup paths found for (*s*, *c*) from relaxed ILP

SP: Set of *k* shortest paths from node *s* to all data centers

α_i^p: Primary path $p \in P$ for (*s*, *c*) passes WMD-zone *i*

β_i^b: Backup path $b \in B$ for (*s*, *c*) passes WMD-zone *i*

PP: Set of already-provisioned primary paths

PB: Set of already-provisioned backup paths

Output:

(*bestP*, *bestB*): primary path *bestP* and backup path *bestB*

```
1.  Compute $S = \{(p,b): max_i(\alpha_i^p \times \beta_i^b) = 0; p \in P, b \in B\}$
2.  if S is not empty and for at least one pair in S, all links on the two paths have enough capacity, then
3.      Take lowest cost pair $(p', b') \in S$ such that all links on the two paths have enough capacity. Set $bestP = p'$ and $bestB = b'$.
4.  else
5.      if P is empty then
6.          GOTO Step 20.
7.      end if
8.      Take lowest-cost primary path $minP \in P$ such that all links on $minP$ have enough capacity. Set $bestP = minP$.
9.      Compute $candidateB = \{b' : b' \in SP, b'$is WMD-zone-disjoint to $bestP\}$.
10.     if candidateB is not empty and for at least one path of candidateB, all links on that path have enough capacity, then
11.         Take least-cost pair $(bestP, b')$ where $b' \in candidateB$ and links on $b'$ have enough capacity. Set best$B = b'$.
12.     else
13.         From the topology graph find (if possible) a shortest path $b'$, WMD-zone-disjoint to bestP, from node S to any of the data center nodes that has content c. Set $bestB = b'$.
14.         if bestB not found then
15.             Delete minP from P. GOTO Step 5.
16.         end if
17.     end if
```

```
            zone-disjoint **then**
                **if** *minCost* > cost($p_1 + p_2$) **then**
                    *bestPrimary*=$p_1$,
                    *bestBackup*=$p_2$.
                    *minCost* = cost($p_1 + p_2$)
                **end if**
            **end if**
        **end for**
    **end if**
**end for**
```

```
**end if**
**if** ($bestP, bestB$) not found **then**
    Use Algorithm 6 to compute *bestPrimary* and
    *bestBackup*. Set *bestP*=*bestPrimary* and
    *bestB*=*bestBackup*
**end if**
```

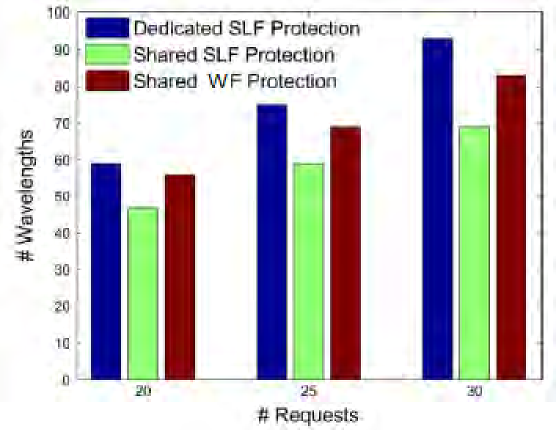


Figure 18. Wavelength usage using Integrated ILP.

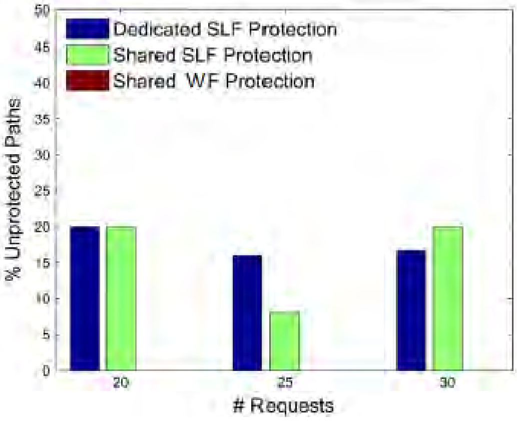


Figure 19. Percentage of unprotected paths using Integrated ILP.

We first compare the wavelength usage of the three schemes. We have three datacenters (located at nodes 2, 6, and 9), 10 contents, and unconstrained number of replicas per content. Datacenter locations are chosen in a way such that at least one of the datacenters is at most two hops away from a node in the network. Figure 18 compares the wavelength usage for shared WF protection with dedicated and shared SLF protection. We see that shared WF protection uses more wavelengths than shared SLF protection but fewer than dedicated SLF protection. Dedicated SLF protection has more probability of being survivable in case of multiple random link failures than shared SLF protection. But, in reality, failures of multiple non-correlated links are quite unlikely. Rather, it is more likely that a set of correlated links/nodes are down simultaneously due to a WMD attack. Though WMD protection provides more protection, it is not a popular choice as it consumes significantly higher resources than protection against a single-link failure. But we find that WMD protection exploiting anycast in a datacenter network consumes moderate resources while providing the required protection.

Figure 19 shows that, without WF protection, a significant number of connections are vulnerable to WMD failures, even though datacenters are distantly located in the network, and primary and backup datacenters are different in all the three cases. With the same primary and backup datacenter, the connections are never protected against destination (datacenter) node failure. These results indicate that shared WF protection, although it uses fewer resources, provides more protection against WMD attacks than dedicated SLF protection.

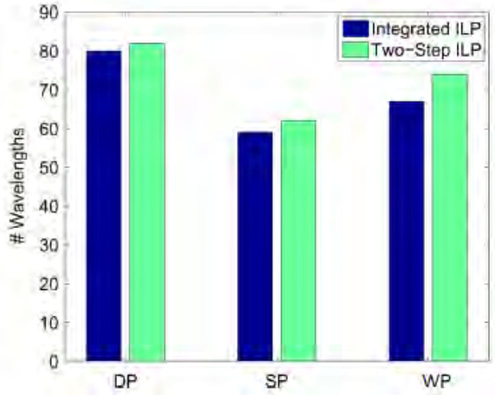


Figure 20. Comparison of two-step ILP with integrated ILP.

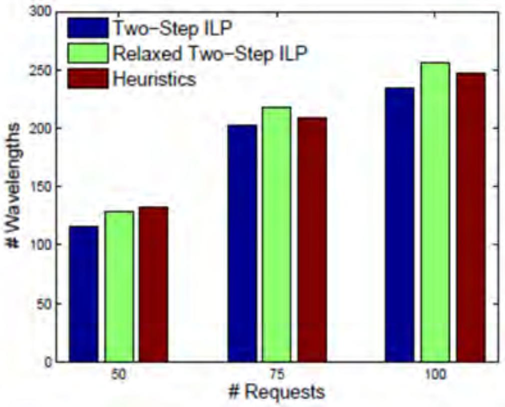


Figure 21. Comparison of two-step ILP, relaxed two-step ILP, and heuristic.

To check the performance of the two-step ILP, we compare its wavelength usage with that of the integrated ILP for 30 connections, 5 datacenters (at nodes 2, 6, 7, 11, and 14), and 3 replicas per content. Figure 20 shows that the performance of the two-step ILP is quite close to that of the integrated ILP. Here, DP, SP, and WP are short forms for dedicated SLF protection, shared SLF protection, and shared WF protection, respectively. For different protection schemes, the wavelength usage in the two-step ILP is 2.5% to 10.4% more than that in the integrated ILP. Note that the result of the two-step ILP is closer to what can be achieved in a real-world scenario where future requests are not known beforehand and replication is done separately from routing. We find that the two-step ILP is more scalable in the number of connections than the integrated ILP. We compared the performance of two-step ILP, relaxed two-step ILP, and heuristics as shown in Fig. 21 and found that their performance is quite close to each other. For more details of these results, we refer to [11].

WMD-Aware Datacenter (DC) and Content Placement

With the rise in demand for cloud services, it is crucial that the network supporting such services be resilient to data loss or service disruptions. Usually, contents are replicated in multiple datacenters (DCs) to ensure availability, but failures due to WMD attacks can make the contents vulnerable to loss. Our study addresses a design problem for WMD-aware datacenter and content placement in a cloud network. We devised a risk analysis to estimate the probabilistic damage such as expected content loss of a given network setting in case of possible attack scenarios and define it as *risk* [12]. Such risk evaluation can be applied to any given existing DC and content placement setting to understand the possibility of content loss due to a WMD attack.

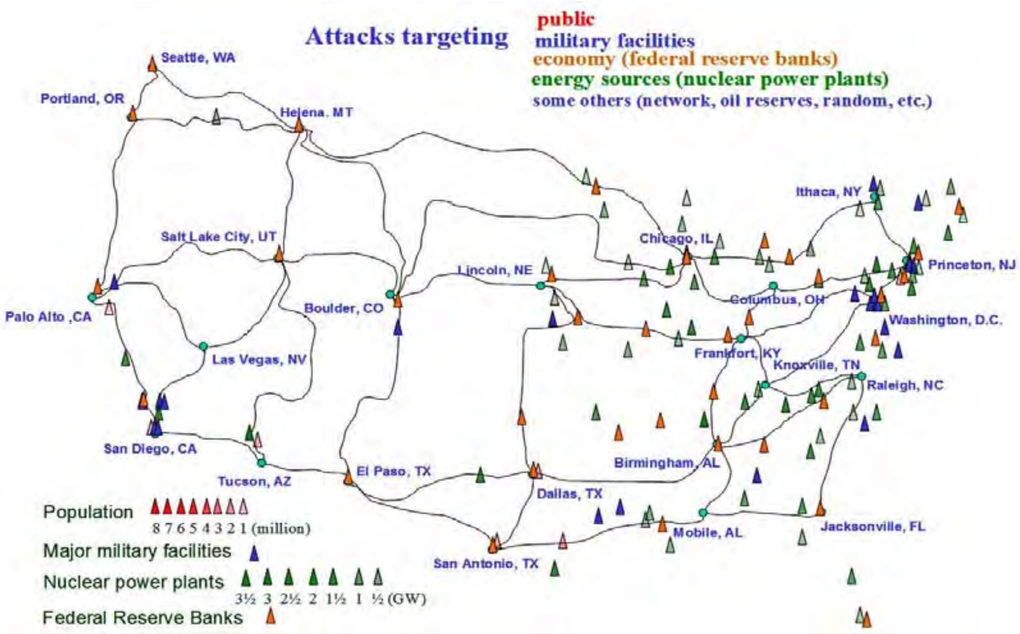


Figure 22. Risk map for WMD attacks on US-wide topology.

We consider two contributing factors to the expected content loss: *(a) Loss due to unavailability:* A content is vulnerable to loss if a DC hosting the content is unavailable due to an attack, i.e., if a DC hosts a content, and if that DC is damaged and is unavailable due to an attack, then the content also becomes unavailable. *(b) Loss due to unreachability:* A content is vulnerable to loss if a DC hosting the content is unreachable from a user node due to a WMD attack. Thus, if a DC hosts a content and if the routes connecting the users to the DC are damaged due to an attack, then the content becomes unreachable from the users.

Here, the reachability of a content is similar to content connectivity in [10]. We investigate the expected content loss due to unreachability with a reasonable finite number of link-disjoint available routes between the user/requesting node and a DC considering latency requirements. While finding the optimal locations for placing a DC, we ensure that the DCs are not placed too far away from the users. A user can obtain a content from a DC via one of the link-disjoint *k*-shortest paths between the user node and the DC. We introduce, for each content, an importance factor which represents the value of the content. For mission-critical (e.g., military) applications, the importance is a more appropriate choice than the traditional content popularity metric. By using this factor, we ensure the survivability of the most important content first. We do not consider full replication where every content is replicated at every DC because this may create more background traffic between the DCs due to synchronization. We estimate the number of replicas of a content as the weighted sum of its importance factor and demands; and choice of weights can depend on the service provider's requirements. We analyze the network with a probabilistic disaster model. A network equipment in a WMD zone (represented as a shared risk group (SRG), i.e., set of links and nodes that fail simultaneously) fails with some probability, which depends on the dimensions of equipment (e.g., link length), its distance from the attack's epicenter, type of attacks, etc. [3]. Probability of a DC or a path (connecting a user node and a DC) being damaged due to an attack is obtained as inputs, based on which the probability of unavailability and unreachability is determined.

Given a network topology, a set of user nodes, a set of candidate DC locations, a set of contents (with corresponding demands, importance factors, and maximum number of replicas), a set of pre-computed link-disjoint *k*-shortest paths between the nodes, and a set of possible WMD attack events with given failure/damage probabilities, we have to assign DCs and contents to candidate DC locations such that the risk (*expected content loss*) of the network is minimized. We formulate the problem as an Integer Linear Program (ILP) with the objective of risk minimization [12].

Input Parameters:

- $G(V,E)$: Physical topology of the network; V is the set of nodes and E is the set of links.
- $D \in V$: Set of candidate DC locations.
- $S \in V$: Set of source (requesting) nodes.
- M: Set of WMD attack events.
- L_{sd}: Set of paths between source node $s \in S$ and DC node $d \in D$.
- C: Set of contents.
- η_c: Demand in terms of number of user requests for content$c \in C$.
- R_c: Maximum number of replicas per content $c \in C$.
- β: Number of DCs to be placed.
- α_c: Importance metric of content $c \in C$.
- $X_d^m \in \{0,1\}$: 1 if node $d \in D$ is damaged by attack $m \in M$.
- $Y_{kds}^m \in \{0,1\}$: 1 if path $k \in L_{sd}$ is damaged by attack r $m \in M$.
- P_{ds}^m: Probability of unavailability or unreachability of $\text{node}\, d \in D$ from node $s \in S$ due to attack $m \in M$.
- P_d^m: Probability that node $d \in D$ is damaged by attack $m \in M$.
- Q_{kds}^m: Probability that path $k \in L_{sd}$ to node $d \in D$ from node $s \in S$ is unreachable due to attack $m \in M$.
- $G_{cs} \in \{0,1\}$: 1 if node $s \in S$ requests content $c \in C$.

Variables:

- $B_d \in \{0,1\}$: 1 if there is a DC at node $d \in D$.
- $T_{dc} \in \{0,1\}$: 1 if content $c \in C$ is available at node $d \in D$.
- $A_{dc}^m \in \{0,1\}$: 1 if content $c \in C$ is unavailable at node $d \in D$ due to attack $m \in M$.
- $H_{dcs}^m \in \{0,1\}$: 1 if node $d \in D$ containing content $c \in C$ is unreachable from node $s \in S$ due to attack $m \in M$.
- $J_{dcs}^m \in \{0,1\}$: 1 if content $c \in C$ in node $d \in D$ is lost from node $s \in S$ due to unavailability or unreachability caused by attack $m \in M$.

Objective:

$$Minimize\colon \sum_{m \in M} \sum_{c \in C} \sum_{s \in S} \sum_{d \in D} \alpha_c \,.\, J_{dcs}^m \,.\, P_{ds}^m \qquad (35)$$

Constraints:

$$J_{dcs}^m \geq \frac{A_{dc}^m + H_{dcs}^m}{2} \qquad \forall s, \forall d, \forall c, \forall m \qquad (36)$$

$$J_{dcs}^m \leq A_{dc}^m + H_{dcs}^m \qquad \forall s, \forall d, \forall c, \forall m \qquad (37)$$

$$A_{dc}^m = X_d^m . T_{dc} \qquad \forall s, \forall d, \forall m \qquad (38)$$

$$H_{dcs}^m = \prod_{k \in L_{sd}} Y_{kds}^m \,.\, T_{dc} \qquad \forall c, \forall m \qquad (39)$$

$$B_d \geq \frac{\sum_{c \in C} T_{dc}}{Z} \qquad \forall d \qquad (40)$$

$$B_d \leq \sum_{c \in C} T_{dc} \qquad \forall d \qquad (41)$$

$$\sum_{d \in D} T_{dc} \leq R_c \qquad \forall c \qquad (42)$$

$$\sum_{d \in D} T_{dc} \geq 2 \qquad \forall c \qquad (43)$$

$$\sum_{d \in D} B_d \leq \beta \qquad (44)$$

$$J_{dcs}^m \geq G_{cs} \qquad \forall s, \forall d, \forall c, \forall m \qquad (45)$$

The objective function in Eqn. (35) computes the overall risk of the network in terms of expected content loss based on α_c due to both unavailability and unreachability. Eqns. (36) and (37) indicate that a content can be lost due to unavailability or unreachability or both. Eqns. (38) and (39) are required to determine content loss due to unavailability and unreachability, respectively. Eqns. (40) and (41) indicates whether node *d* hosts a DC (here, Z is a large integer constant). Eqns. (42) and (43) provide a bound on the number of replicas of content *c*. Eqn. (44) gives a bound on the total number of DCs in the network. To ensure content availability, there should be at least two replicas of a content *c*, whose upper bound is R_c. Eqn. (45) indicates that a source node *s* obtains a content *c* if *s* has demand for *c*, i.e., we consider unavailability and unreachability of *c* at *d* from *s* due to attack *m* only when *s* requests for *c*.

We also explore a WMD-unaware datacenter and content placement design to evaluate the effectiveness of our WMD-aware approach. The objective of WMD-unaware placement is to minimize the average path-cost, (i.e., average distance from a user node to the nearest replica of a content), while placing the DCs. The path distance is proportional to the access latency.

To evaluate the benefits of our design, we simulated a 24-node USnet topology, with a focus on WMD attacks. Based on Fig. 22, we consider possible locations of major military facilities that are probable targets of WMD attacks and modeled 10 WMD attack zones as shown in Fig. 23. We consider not only primary attacks but also correlated/cascading effects such as secondary attacks and power outages. Based on information in [13] and considering multiple correlated/cascading effects, we assume a failure span of 1000 km around the targeted areas. The probabilities of damage/failure on nearby nodes and links are estimated with reasonable assumptions based on their distances from the target's epicenter. For example, in Fig. 23, due to attack event m_1 (shown with corresponding WMD zone), two links (3-4 and 3-7) are estimated to be damaged with probability 1; and nodes 3, 4, and 7 have estimated damage probabilities of 0.7, 0.4, and 0.05, respectively (decreasing with distance). Similarly, due to m_2, node 5 is estimated to be damaged with probability 1; hence, all associated links are failed, and neighboring nodes 3 and 4 have estimated damage probabilities of 0.05 and 0.1, respectively. We consider all network nodes as candidates for DC and we consider eight DCs. The user nodes are distributed according to population density and their vicinity to military facilities because these regions will generate the most number of requests, and we consider all nodes as user nodes. Typically, in a cloud network, the number of contents can be very large but, since the ILP has limited scalability, we first consider a small number of contents, e.g., 20, with some assigned importance factors (on a scale of 1-10). We assume all user nodes can request all the contents.

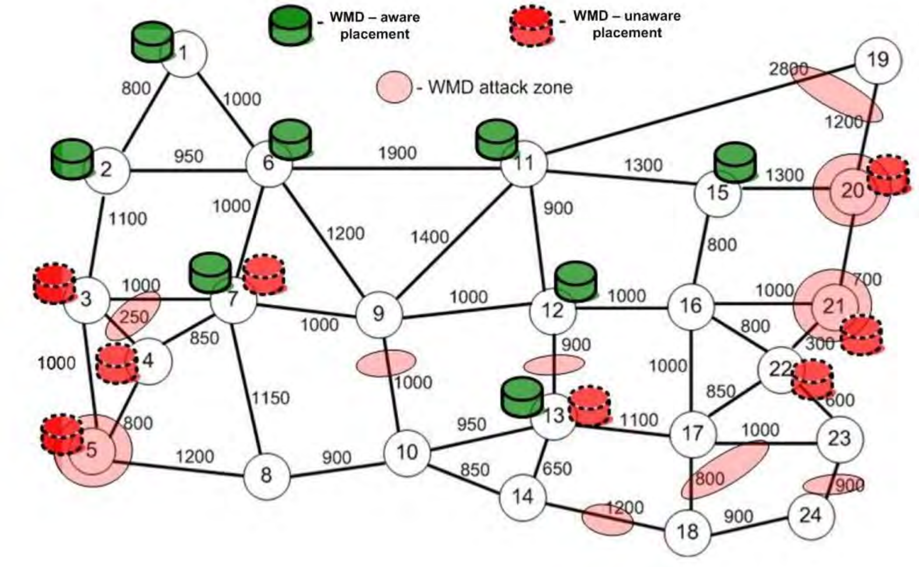


Figure 23. WMD-aware and WMD-unaware datacenter and content placement based on risk map.

We compare WMD-aware and WMD-unaware datacenter and content placement approaches. As shown in Fig. 23, WMD-aware placement (shown in bold lines) assigns datacenters in locations such that nodes vulnerable to WMD attacks are avoided. For instance, based on attack events, worst candidate DC locations are nodes 5, 20, 21, and 19; and next worst candidates are nodes 3 and 4 since they are associated with high link failures from both attack events m_1 and m_2. The contents are distributed according to their importance and their demands from user nodes. The WMD-unaware placement (shown in dashed lines) assigns datacenters to locations that may be vulnerable to attacks since now the objective is only to minimize the path-cost from user nodes to DCs. In case of an attack, DCs at nodes 5, 20, and 21 will incur huge expected content loss due to unavailability whereas DCs at nodes 3 and 4 will incur high expected content loss due to unreachability. Such placement may not be desirable.

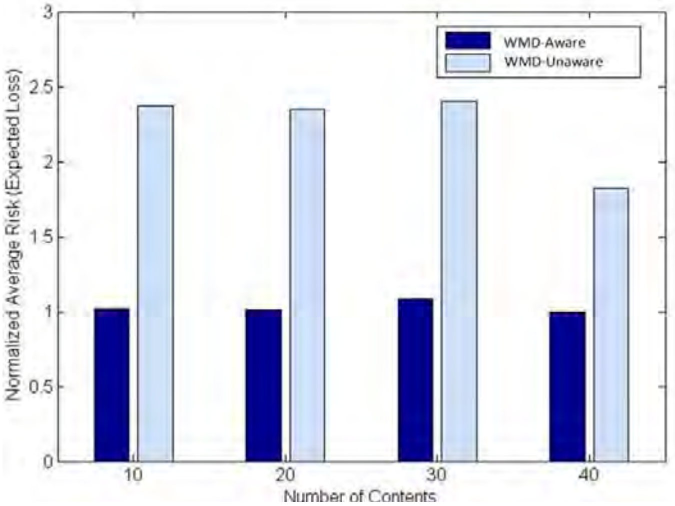


Figure 24. Risk comparison of WMD-aware and WMD-unaware approaches.

We compare the expected content loss (i.e., risk) of the two approaches in Fig. 24. The WMD-aware approach shows significant improvement over WMD-unaware approach (45-57% risk reduction for the number of contents ranging from 10-40) for this case study. We also showed in [12] that our WMD-aware approach is not greatly resource-intensive (in terms of average path-cost.) compared to WMD-unaware approach.

WMD-Aware Dynamic Content Placement

The WMD-aware datacenter and content placement approach gives us locations where DCs and contents are placed with minimum risk based on some attack probabilities. But content properties (user demands and importance) and probability of WMD attacks can be time-varying. With updated network information, content placement should be such that the expected content loss is curtailed at any given time. A new attack scenario can make a currently-safe DC location a risky location at a later time. Once established, it is not viable to relocate a DC depending on changing risk profile. Rather, with a fixed setting of DCs, there should be a rearrangement mechanism so that if a content requires more replicas or requires a safer placement, it can be re-replicated at new locations. But rearranging contents among a set of DCs is resource-intensive since it involves adding/deleting replicas. There should be a way to reduce the number of rearrangements and hence replication cost, e.g., network bandwidth consumption (cost is linearly proportional to number of replicas), and background traffic for synchronization.

We propose a WMD-aware dynamic content placement algorithm [12] which decides where a content is to be re-replicated so that risk or *expected content loss* of the network is reduced under dynamic settings of demands and WMD attack events. While performing content placement, we consider users' QoS metrics, such as *latency*, i.e., we want to place contents as close as possible to users. Satisfying latency constraint with risk reduction is a major challenge because these two objectives can be contradictory. Placing contents in less-risky locations can reduce risk, but such locations may be distant from users and vice versa. We consider an average latency-aware distance which is determined by QoS. If QoS cannot be satisfied due to risk reduction requirements, users will at least have the content available via a longer path. Also, we try to reduce the number of replicas by deleting redundant replicas to reduce resource usage.

We consider that service providers will employ a reasonable *threshold risk* value per content which satisfies the expected level of availability for that content against a WMD attack. This value may differ from content to content depending on the importance factor or the provider's requirements. Our goal is to achieve this threshold value, or, in case the threshold requirement cannot be met, at least to achieve the minimum possible risk with given attack alerts. We analyze risk based on the existing DC placement with new or changed settings: new set of user nodes (hence new demands), new required number of replicas of contents (based on new demands and new importance factors), and new WMD attack probabilities.

The algorithm solves the problem with the objective of reducing the risk in terms of expected loss of content, while 1) reducing the number of content rearrangements in the network and 2) satisfying the QoS metric (minimum access latency) of user nodes. We further reduce the number of replicas, if possible, within risk requirements. Given a network with existing set of DC locations and updated settings, our proposed heuristic updates the placement of *one content at a time* for global risk reduction for all contents. We compute at run-time some risk parameters based on Eqn. (35) for *each* content (see [1]) and based on these values, we decide if a content placement should be updated or not. As shown in Fig. 25, the algorithm (see [12]) is divided into five phases, each addressing a separate goal. For each content, we retrieve the set of DCs where the content is initially replicated, and set of remaining DCs which are candidate locations for new placement.

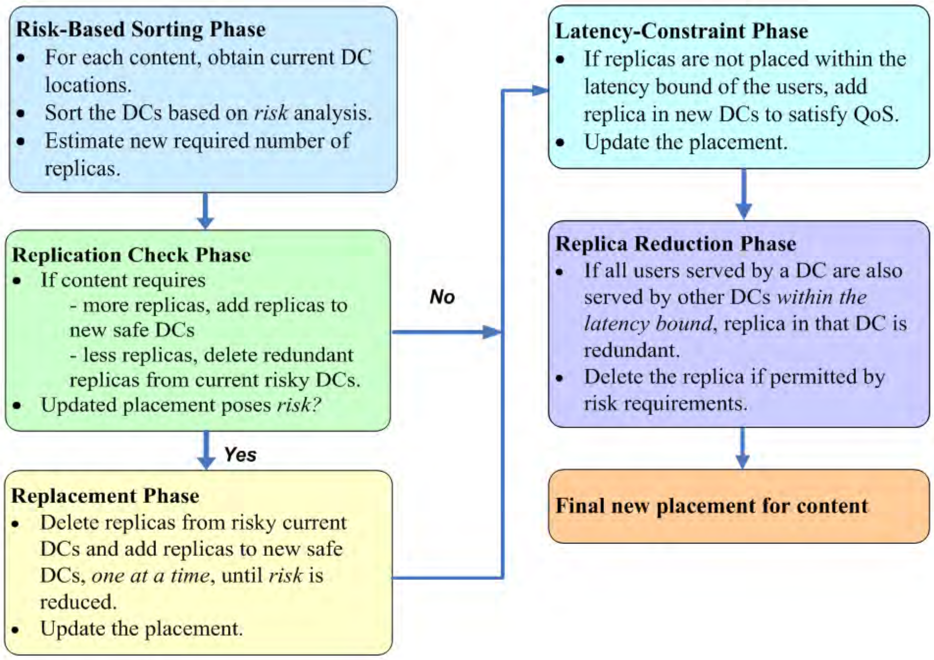

Risk-Based Sorting Phase
For each content, obtain current DC locations.
Sort the DCs based on risk analysis.
Estimate new required number of replicas.
Replication Check Phase
If content requires
- more replicas, add replicas to new safe DCs
- less replicas, delete redundant replicas from current risky DCs.
Updated placement poses risk?
No
Yes
Replacement Phase
Delete replicas from risky current DCs and add replicas to new safe DCs, one at a time, until risk is reduced.
Update the placement.
Latency-Constraint Phase
If replicas are not placed within the latency bound of the users, add replica in new DCs to satisfy QoS.
Update the placement.
Replica Reduction Phase
If all users served by a DC are also served by other DCs within the latency bound, replica in that DC is redundant.
Delete the replica if permitted by risk requirements.
Final new placement for content

Figure 25. WMD-aware dynamic content placement algorithm.

We sort all DCs based on the *partial risk* which they may incur with updated settings, to determine which DC locations are safe (low risk) and which are risky for this content. We also determine the new required number of replicas for the content; and, depending on the requirement, we either add replicas in new safer candidate DCs or we delete redundant replicas from risky DCs. We then check if the current placement of a content achieves the *threshold risk* with the updated settings. If not, we rearrange the placement of one replica at a time and analyze if the new risk is acceptable or not. We start by deleting a replica in the riskiest DC location in the initial placement and adding in the new safest candidate DC. In worst case, all replicas may need to be re-replicated at new safer DC locations, but then, this would achieve the best possible placement or the minimum achievable risk, at the expense of resource usage (e.g., number of wavelengths, bandwidth, etc.) After risk reduction, we re-adjust by adding replicas to satisfy the QoS of latency constraint. We then try to reduce the number of replicas as long as risk requirements and user demands are met. The resultant placement is the final placement for the content. We thus obtain the updated placements of all contents and then determine the total risk in the network.

We simulated a 24-node USnet topology under the same settings as before. We alter/adjust the values of content demands, importance factors, and failure probabilities within a reasonable small range to simulate the dynamic setting. For our examples, we used *threshold risk* = 100 based on some repeated experiments with different inputs for different number of contents. We used *latency distance* = 3000, based on the distances (in km) in the given topology.

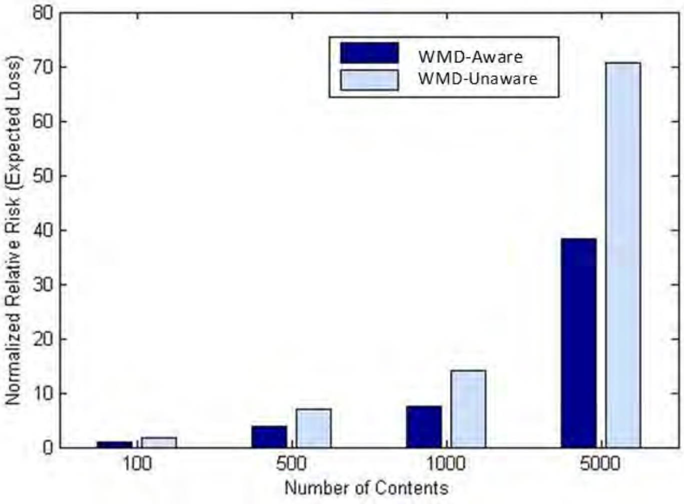


Figure 26. Risk comparison of WMD-aware dynamic content placement.

In Fig. 26, we show expected content loss or *risk* for large number of contents (100-5000). For comparing the WMD-aware approach with a WMD-unaware approach, we employ content placement based on demands in the network without any risk minimization. We then apply our heuristic on the WMD-unaware placement and compare the risk. Our results show that the WMD-aware approach gives significant improvement in terms of risk reduction (up to 45%), for this example. We also see that our approach performs better with increasing number of contents.

Rapid Data Evacuation

Cloud services delivered by datacenter (DC) networks, can range from online data storage and processing, web applications to more critical collaborative applications which have small tolerance to service disruptions and may require data transfer in the range of Terabytes to Petabytes. It is crucial to maintain availability of these services even under extreme circumstances of WMD attacks. A less-occurring, yet more damaging form of WMD attack can be an electromagnetic pulse (EMP) attack since all basic infrastructures depend on the civilian electric grid which is largely unsecured against such sudden and extensive attack. According to the Federation of American Scientists, “A large device detonated at 250 to 312 miles over Kansas would affect all of the continental U.S.” [14]. Cloud services intrinsically yield some degree of protection against attacks in terms of data redundancy as contents/services are typically replicated in geographically distributed DCs. Different replicas in different DCs are synchronized periodically to ensure that all the replicas are updated. In case of very large-scale WMD attacks, disrupting multiple DCs across cities, it may happen that, *all* (or the most updated) replicas of a content are within the attack region and are vulnerable to loss. To save critical data under such circumstances, contents should be moved away or evacuated from affected locations to secure locations. We assume that service providers will receive warning before an oncoming attack. The duration between receiving a warning and the predicted time of the actual attack is very crucial since it is the time interval, in which servers in the DCs can transfer vital data to safe locations before becoming non-operational. Depending on attack scenario, this interval can range from millisecond to a few minutes.

Our study addresses a rapid data-evacuation strategy in an optical cloud network for large-scale WMD attacks. Data evacuation essentially means transferring or replicating contents from an WMD affected location to secure location. Typically, inter-DC data transfers are considered delay tolerant but under large-scale attack scenarios, data has to be transferred under real-time threat of an oncoming attack, which makes available time and resources very limited. Due to strict time and resource constraints enforced in an attack scenario, only the vulnerable contents (contents for which *all* replicas are in attack zone) are considered for evacuation. Each content is associated with an importance metric, to prioritize the contents during evacuation. Usually, DCs are equipped with huge storage capacity and they can allow large inter-DC data transfers, but in case of large-scale attacks, large data evacuation from affected DCs can limit the storage. Destination site(s) for evacuated data are chosen outside attack zone based on distance (for quick transfer) and available storage. We consider inter-DC network with a fiber-optic backbone that provides high-bandwidth dynamic circuit-switched paths. But even with high capacity, network can be highly utilized due to increased traffic in response to attack threat. Hence, data evacuation is largely constrained by time, network bandwidth (residual link capacity), and DC storage resources. We propose a rapid data-evacuation strategy [15] which identifies the endangered contents to be evacuated from a WMD affected region, and with the available network bandwidth and storage capacities of secure DCs, finds the least delay paths for data evacuation. Our goal is to maximize the amount of contents evacuated under the *evacuation deadline* constraint which is determined by the type and propagation of attack. Given a network topology, set of pre-computed *k*-shortest paths, set of possible WMD attack zones, set of DC nodes, with storage capacity and a set of locally hosted contents with corresponding number of replicas, importance metric, α_c and size, residual link capacity, and evacuation deadline, find the least delay paths for data evacuation.

Depending on the attack, service providers set an *evacuation deadline*, *T,* during which vulnerable contents should be evacuated. We compute data transfer time of considering the delays incurred during *path computation*, *switch configuration*, and *data transfer*. We model the timeline of a data-evacuation process as follows: *Threat received → initiate evacuation strategy → path computation → connection setup → transmission → propagation.* Path computation delay depends on the computation time of the evacuation algorithm. Connection setup delay includes switch configuration delay, control message processing, and propagation delay. We assume that the control plane is reliable and that transmission delays for control messages are negligible; connection setup delay model is used as in [5]. While other delays are typically in millisecond range, transmission delay is the most significant delay which depends on the channel data rate and data volume. Since an attack imposes time-sensitive threats, rapid data-evacuation strategy schedules evacuation of each content in quickest possible time, and hence, uses full available bandwidth for a path during data transfer. As a result, the resources of the path become occupied and a subsequent content can be scheduled only after the previous transfer is complete. For each subsequent content, we check if the computed path shares a link with any path used for transferring prior contents. If there is link sharing, then previous transfers on that path has to be completed and hence, delays incurred by previous contents are acquired in the data transfer delay of the current content. On the other hand, if there is no path sharing with any previous transfer, our strategy can schedule the content transfer independently, and no previous transfer delay will be acquired. Hence, for each content, data transfer delay on a path is computed as: {path computation delay + connection setup delay + transmission delay + propagation delay + acquired delay (if any)}. The path that gives the least overall delay and satisfies the *evacuation deadline* constraint is setup for evacuation of the content. The heuristic for rapid data evacuation is described in Algorithm RDES [15].

Algorithm: Rapid Data-Evacuation Strategy (RDES)

Input: Network topology G(V,E), set of pre-computed k-shortest paths L, set of possible attack zones M, set of DC nodes $\mathrm{D} \in \mathrm{V}$ with storage capacity $\mathrm{S_d}$, $\mathrm{d} \in \mathrm{D}$ and a set of locally hosted contents $\mathrm{C_d}$, $\mathrm{d} \in \mathrm{D}$ with corresponding number of replicas $\mathrm{R_c}$, importance metric α_c and size $\mathrm{F_c}$, residual link capacity $\mathrm{B_e}$, $\mathrm{e} \in \mathrm{E}$, and evacuation deadline, T.

Output: Amount of contents evacuated by T.

- <u>Mapping of attacks:</u> Find set of DCs $\mathrm{D_M} \in \mathrm{D}$ within WMD attack zone $\mathrm{m} \in \mathrm{M}$ and set of DCs $\mathrm{D_{notM}} \in \mathrm{D}$ outside attack zone m.

- <u>Content selection:</u> Find set of contents, $\mathrm{C_{Evac}} \in \mathrm{C}$ for which all replicas are hosted in DCs inside attack zone m. Sort $\mathrm{C_{Evac}}$ based on α_c in descending order.
- <u>Destination DC selection and path computation for each content:</u>
    - For each content $\mathrm{c} \in \mathrm{C_{Evac}}$,
        - Obtain $\mathrm{D_c}$, set of DCs which host replicas of c. Find set of candidate DCs, $\mathrm{DC_{cand}}$ which are outside attack zone m and which satisfy the storage constraint ($\mathrm{F_c} \leq \mathrm{S_d}$) for c.
        - Compute set of k-shortest paths, L between every DC $\mathrm{d_{src}} \in \mathrm{D_c}$ and every DC $\mathrm{d_{dest}} \in \mathrm{DC_{cand}}$.
            - For each path $\mathrm{p} \in \mathrm{L}$, identify the bottleneck link $\mathrm{e} \in \mathrm{E}$ (link with lowest residual capacity along that path) and set $\mathrm{B_e}$ as the bandwidth of path p. Compute data transfer delay, $\mathrm{delay_{tx}}$ for c on p.
            - If p shares a link with any path that has already been computed for transferring any prior content, add $\mathrm{delay_{acquired}}$ (maximum data transfer delay incurred by any prior content on the path); $\mathrm{delay_{total}} = \mathrm{delay_{tx}} + \mathrm{delay_{acquired}}$. Else, $\mathrm{delay_{total}} = \mathrm{delay_{tx}}$.
        - If path with minimum $\mathrm{delay_{total}}$ satisfies ($\mathrm{delay_{total}} \leq \mathrm{T}$), the path is evacuation path for c between DC $\mathrm{d_{src}}$ and DC $\mathrm{d_{dest}}$. Else, no evacuation path exists.
        - Update storage capacity of DC $\mathrm{d_{dest}}$ accordingly. Update $\mathrm{D_c}$.
    - Setup connections based on the computed evacuation paths for transferring data.

To evaluate the benefits of our design, we simulated our heuristic with a 8-node DC network mapped on a 24-node USnet topology with a focus on EMP attacks as shown in Fig. 27. Nodes with thick black circles represent DC nodes. The areas in different shades of red represent different span and intensity of EMP attack [14]. For our simulation, we consider areas covered by the deepest and medium shades of red as attack zone in which nodes and links fail with probability 1. DCs at nodes 6, 9, and 12 are directly under the threat of the attack. The 8 DCs have free storage capacity ranging from 10 TB to 100 TB, and every DC is connected with high-bandwidth links with residual capacity ranging from 500 Gbps to 1 Tbps. The network is assumed to have 30% utilization, carrying regular inter-DC traffic. The number of contents = 100, and average size of contents ranges from 20 GB to 100 GB; contents are modeled as cumulative of multiple smaller contents. The contents are uniformly distributed among DCs with number of replicas ranging from 2 to 4. Each content is assigned α_c on a scale of 1-10 using an uniform distribution. For our simulations, we used processing delay = 10 μs, propagation delay = 5 μs/km, and switch configuration delay = 15 ms [5]. In our simulation, around 35% of the total contents were chosen for evacuation. For comparing our rapid data-evacuation approach, we employ nearest evacuation approach which evacuates data only to nearest DC with shortest path.

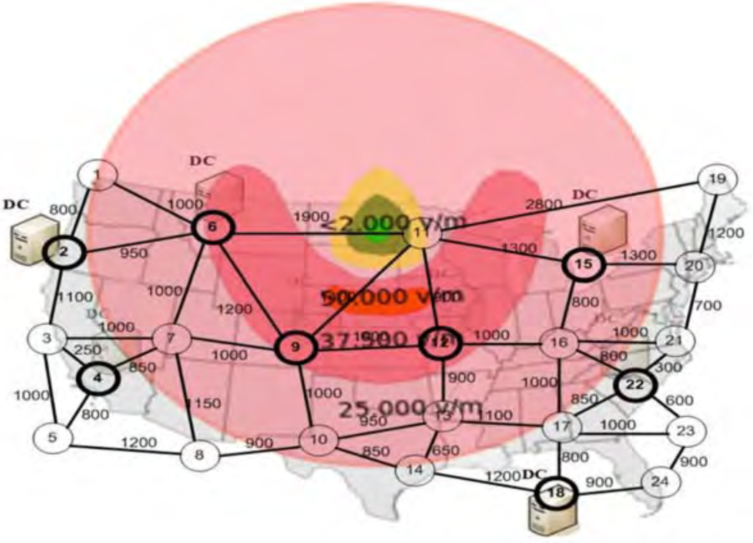


Figure 27. Span of an EMP attack on US-wide topology.

In Fig. 28, we show cumulative volume of contents evacuated, using the two approaches, according to *evacuation deadline*, *T*. Our results show that, as time constraint is relaxed, more contents are evacuated and our rapid-evacuation approach evacuates data much faster than the nearest evacuation approach. It takes about 19 sec to evacuate all data with the nearest evacuation approach, whereas our approach with intelligent path computation takes about 7 sec (63% less) which is a significant time savings in case of attacks.

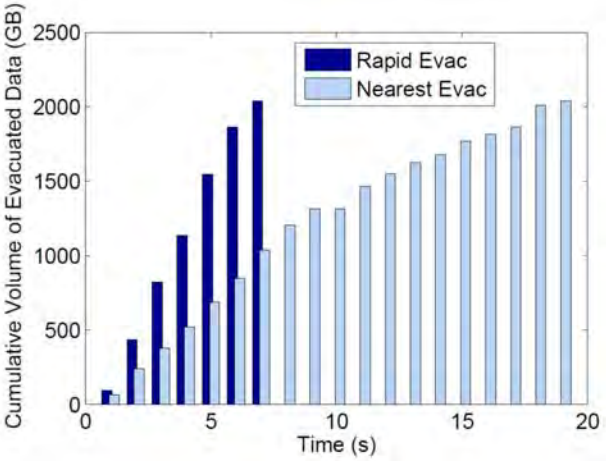


Figure 28: Comparison of Rapid Evacuation with Nearest Evacuation.

Under normal mode of operation, the network has some traffic utilization and hence all network resources are not fully available. But during an emergency large-scale attack scenario, priority can be given to data evacuation over existing connections. In Fig. 29, we show the effect of increasing the residual link capacity by 10% - 50% on the volume of data evacuated with *T* = 4 sec. From Fig. 28, we can see that around 1130 GB volume of data were evacuated by T = 4 sec. Figure 29 shows that, with increased capacity, significant volume of additional data can be evacuated, and with 50% capacity increase, almost all the scheduled contents (total volume of around 1800 GB) could be evacuated within T = 4 sec. Drawing from our result in Fig. 28, evacuation of such volume of contents required 6 sec. Time savings of such scale can be very crucial in case of large-scale attacks.

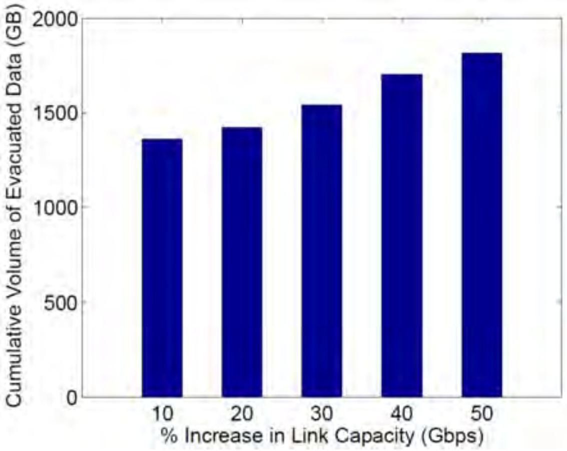


Figure 29: Rapid Data Evacuation in T = 4 sec with Increased Capacity.

Disaster/Post-Disaster Survivable Cloud Network Mapping with Risk Minimization

Cloud-network (CN) mapping is the combination of virtual-network (VN) mapping and virtual-machines (VMs) allocation (i.e., network and server virtualization) over a physical infrastructure. CN survivability is crucial for computational resource allocation in a consistent and secure environment for cloud-computing services, e.g., the Department of Homeland Security (DHS) is building its Private Cloud to manage sensitive information which may use one or several cloud networks (CNs) [9]. Failures in the physical infrastructure can reduce the available resources (optical network and datacenter) and disconnect multiple CNs, endangering the cloud data and services. Given the scale of the impact of WMD/disaster failures in CNs, network operators should take measures to protect cloud-computing services from WMD attacks and post-attack failures despite their rare occurrences. In the rest of this section, we use WMD attack and disaster interchangeably.

To reduce the disconnection of CN in a WMD attack and post-attack effects, we proposed the Disaster and Post-Disaster Survivable CN mapping with Risk Minimization (Min-Risk-D-PDS) approach, which includes two novel approaches: VM Backup Location (VBL) and Post-Disaster Survivability (PDS) [16].

Virtual Backup Node for VM Backup Location (VBL)

A virtual backup node is necessary to relocate or replicate the affected VMs into safer locations (i.e., datacenters). Min-Risk-D-PDS uses VM Backup Location (VBL) functionality to enable VM replication or relocation in case of WMD attacks. VBL maps one or more virtual-backup node to relocate VMs of a CN, using three main steps: selection, connection, and sharing. The physical node (i.e., datacenter) selected as backup must not only have enough excess processing capacity but also should be located in a safer place to lower the risk of disconnection.

Every virtual backup node has to be connected using one virtual link to a set of working VMs in its own CN (Fig. 30(b)). The selected physical node for VM backup location of one CN can be shared by another CN. For survivability to post-disaster failures, this approach will not allow to share the same physical node if both CNs can be disconnected by the same disaster. Figure 31 presents an example using VBL.

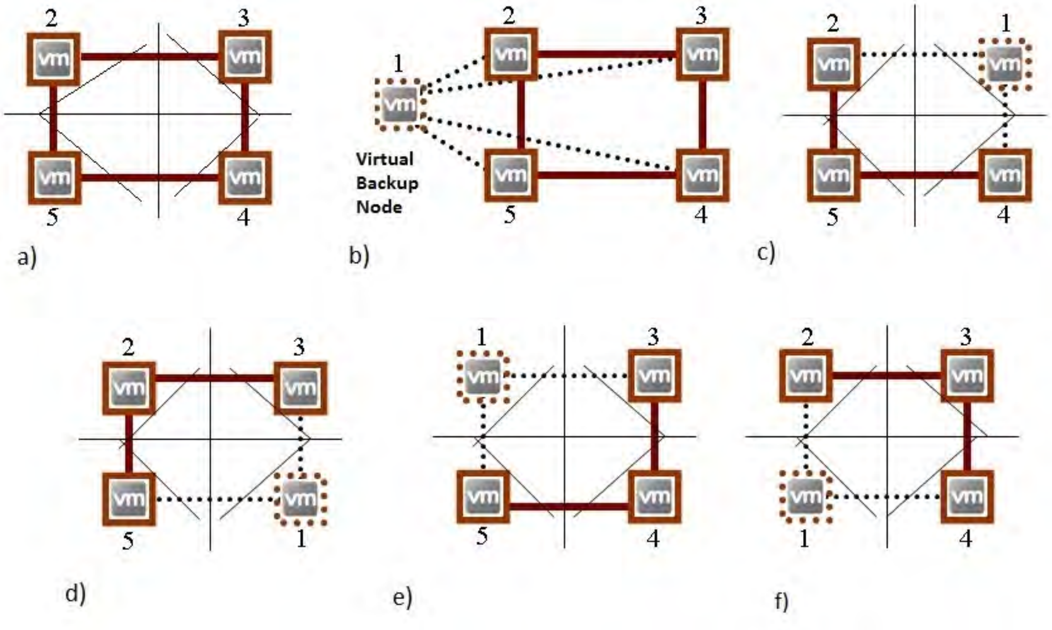


Figure 30. Example of cut extension for post-disaster survivability (PDS). The extended cuts are cuts in the resulting virtual topology when one VM is replaced for a backup VM. a) CN with basic cuts, b) Mapping of the CN with virtual backup node VM 1. c) Extended cuts for the case when VM 3 fails and is replaced by backup VM 1. d) Extended cuts for the case when VM 4 is replaced by backup VM 1. e) Extended cuts for the case of VM 2 is replaced by backup VM 1. f) Extended cuts for the case when VM 5 is replaced by backup VM 1.

Post-Disaster Survivability (PDS)

Post-attack (i.e., post-disaster) correlated and cascading failures may damage additional physical links and/or datacenters not affected by the primary attack. These failures may disconnect CNs and interfere in the recovery process of failed CNs and endanger important information and services. Hence, we add a Post-Disaster Survivability (PDS) approach in our Min-Risk-D-PDS model to increase the survivability during recovery periods. Our (PDS) approach consists of two functions: i) The cut extension functionality gets the basic cuts (Fig. 30(a)) and assumes the possible failure and replacement of any VM as we show in Figs. 30(c), (d), (e) and (f)). ii) A survivability constraint is put in place to handle the extended cuts which consider the virtual connectivity in any VM replacement case. Figure 31(b) shows an example of CN mapping using Min-Risk-D-PDS with VBL and PDS functionalities.

Example of Min-Risk-D-PDS Approach

Figure 31(a) shows a CN network mapping of Min-Risk-D-PDS (without PDS approach). In case of a WMD attack, e.g., in DZ1, the physical node C and its physical links will fail, but the CN will not be disconnected, because the failed VM in node 2 will be relocated into physical node A (VM in node 1). However, a post-disaster failure in physical link A-B will disconnect the CN, because virtual links 1-5 and 1-4 will be disconnected. Similarly, failure of any of physical links B-E, F-G, and E-G may disconnect the CN too. Adding the PDS approach, we obtain the mapping of Fig. 31(b), where the CN will not be disconnected by any single physical-link failure, disaster failure, or post-disaster single physical-link failure, and the expected loss of bandwidth and processing capacity will be reduced.

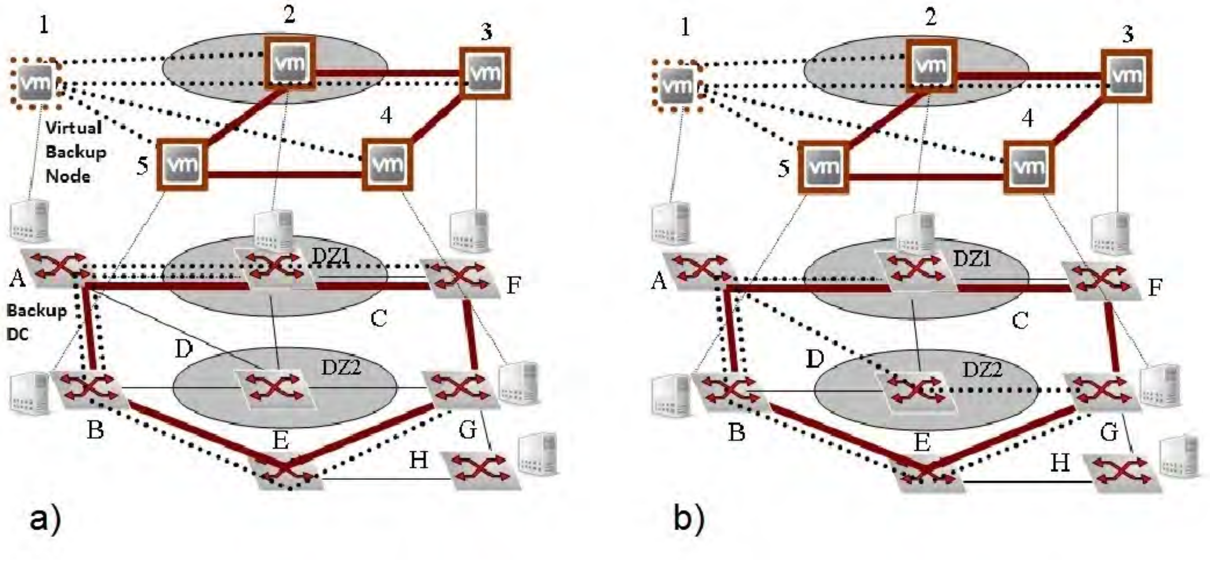


Figure 31. (a) Disaster resilient mapping and (b) Disaster and post-disaster survivable CN mapping.

Mathematical Model

Min-Risk-D-PDS approach is modeled as an integer linear program (ILP) in this section.

Given:

G(*V, E*): Physical topology; *V* is the set of nodes and *E* is the set of links

$\hat{V}$: Set of VM datacenter locations $\hat{V} \subset V$.

$G_{\gamma}(V_{\gamma}, E_{\lambda})$: Topology of CN *γ* where V_{γ} is the set of working VM locations (virtual nodes, $V_{\gamma} \subset \hat{V}$), and E_{γ} set of virtual links of CN.

C_{γ} : Set of basic cuts of CN topology *γ*.

$\hat{E}_{\gamma}$: Set of virtual links including the links in E_{γ} and virtual links from each node in V_{γ} to each node in $\{\hat{V} - V_{\gamma}\}$.

$\hat{C}_{\gamma}$: Set of extended cuts of CN topology *γ* formed by a possible relocation of working VM of V_{γ} to a physical node *b* with free processing capacity in $\{\hat{V} - V_{\gamma}\}$.

$\Gamma = \{\gamma =< V_{\gamma}, E_{\gamma}, C_{\gamma}, \hat{E}_{\gamma}, \hat{C}_{\gamma}\}$: Set of cloud networks (CNs).

$s_{i,j}^{n}$: 1 if the physical link *{i,j}* is disconnected by WMD attack *n*, zero otherwise.

$S_n : \{s_{i,j}^{n}\}, S_n \subset E.$

p_n : Probability of occurrence of WMD attack *n.*

$N = \{< S_n, p_n >\}$: Set of WMD attack zones (i.e., DZs).

P_u^{γ} : Processing capacity required to allocate VM *u* used by CN *γ* ($u \in V_{\gamma}$).

$F_{i,j}$: Capacity of physical link (*i,j*).

P_{free}^{v} : Excess processing capacity in physical node *v.*

d: CN disconnection coefficient (1 ≥ d ≤ 10).

b_e : Bandwidth requirement of virtual link *e.*

b_c : Total capacity that can be lost if the links of the cut *c* are disconnected (i.e., the CN is disconnected).

m_c : Number of virtual links in cut *c*

Binary variables:

D_e^n : 1 if virtual link *e* is disconnected by WMD attack *n*.

$M_{i,j}^{e}$:1 if virtual link *e* is mapped on physical link (i.j).

$K_{u,v}^{\gamma,e}$: 1 if virtual Link e from node *u* to *v* in *γ*.

Y_b^{γ} : 1 if b is assigned as a virtual backup node of *γ*.

Q_c^n : 1 if virtual links of cut c is disconnected by WMD attack *n*.

X_{γ}^{n} : 1 if CN *γ* may be disconnected by WMD attack *n*.

$T_{g,h}^{n}$: Is an auxiliary variable.

$Z_{u,b}^{n}$: 1 if VM *u* can be relocated to datacenter *b,* $b \in \hat{V}$ in case of WMD attack *n*.

Objective Function:

The objective is to minimize the total capacity that can be lost if a WMD attack occurs. The risk is defined as total penalty for capacity loss multiplied by the probability of occurrence (i.e., WMD attack).

The total penalty for capacity lost is the sum of penalty for CN and virtual links' disconnections. The penalty for CN disconnection is calculated by $\sum_{c\in C_\gamma} dQ_c^n b_c$ which is the sum of capacity *b*$_c$ that is lost if a CN is disconnected by WMD attack *n* multiplied by a CN disconnection coefficient *d.* The penalty for virtual-link disconnection is calculated by $\sum_{(i,j)\in E} D_e^n b_e$ which is the sum of capacity *be* that is lost when virtual links e is disconnected by WMD attack *n*. Finally, the objective function is:

$$\min \sum_{n\in N} \sum_{\gamma\in\Gamma} \left(\sum_{c\in C_\gamma} dQ_c^n b_c + \sum_{(i,j)\in E} D_e^n b_e \right) p_n + \left(\varepsilon + \sum_{(i,j)\in E} \sum_{\gamma\in\Gamma} \sum_{e\in\hat{E}_\gamma} M_{i,j}^{e} \times b_e \right) \tag{46}$$

To avoid the mapping of virtual links over long optical paths, a resource-minimization formula is added.

Constraints:

CN and virtual link disconnection constraints in case of a WMD attack:

$$D_e^n \ge \frac{1}{M} \sum_{(i,j)\in E} s_{i,j}^n M_{i,j}^{e} \text{ , } D_e^n \le \sum_{(i,j)\in E} s_{i,j}^n M_{i,j}^{e}, \forall e \in \hat{E}_\gamma, \gamma \in \Gamma, n \in N \tag{47}$$

$$Q_c^n \le \frac{1}{m_c} \sum_{e\in E_c} D_e^n \text{ , } Q_c^n \ge \sum_{e\in E_c} D_e^n - m_c + 1, \forall c \in C_\gamma, \gamma \in \Gamma, n \in N \tag{48}$$

CN mapping and physical links capacity constraints:

$$K_{u,v}^{\gamma,e} = 1, \forall u, v \in V_\gamma, u \ne v, \gamma \in \Gamma, e \in \hat{E} \text{ , } \sum_{e\in\hat{E}_\gamma} M_{i,j}^{e} \le F_{i,j}, \tag{49}$$

Flow-conservation constraints:

$$\begin{aligned} &\sum_{(i,s_e)\in E} M_{i,s_e}^{e} - \sum_{(s_e,j)\in E} M_{s_e,j}^{e} = -K_{s_e,d_e}^{\gamma,e} \sum_{(i,d_e)\in E} M_{i,d_e}^{e} - \sum_{(s_e,j)\in E} M_{d_e,j}^{e} = K_{s_e,d_e}^{\gamma,e} \\ &\sum_{(k,j)\in E} M_{k,j}^{e} - \sum_{(i,k)\in E} M_{i,k}^{e} = 0, \forall e \in \hat{E}_\lambda, \gamma \in \Gamma, k \in \hat{V} - \{s_e, d_e\} \end{aligned} \tag{50}$$

Survivability constraint which ensures resiliency in case of WMD attack disconnect any physical link. This constraint implements the Post-Disaster Survivability (i.e., Post-Attack Survivability) by using the extended cuts of the remaining topology after any WMD attack:

$$\sum_{e\in\hat{E}_\gamma} M_{i,j}^{e} \le m_c - 1, \forall c \in C_\gamma, \gamma \in \Gamma, (i,j) \in E \tag{51}$$

Constraints which ensure that a virtual backup node for VM relocation cannot be shared by CNs that are disconnected by the same WMD attack:

$$X_\gamma^n \ge \frac{1}{M} \sum_{c\in C_\lambda} Q_c^n, \; X_\gamma^n \le \sum_{c\in C_\gamma} Q_c^n, \forall \gamma \in \Gamma, n \in N \tag{52}$$

$$T_{g,h}^{n} \le X_{g}^{n}, T_{g,h}^{n} \le X_{h}^{n}, \quad T_{g,h}^{n} \ge X_{g}^{n} + X_{h}^{n} - 1$$

$$Y_{b}^{g} + Y_{b}^{h} \le 2 - T_{g,h}^{n}, \forall g, h \in \Gamma_{\gamma}, g \ne h, n \in N, b \in \left[\hat{V} - (V_{g} \cup V_{h})\right]$$

Constraint which maps the virtual-backup node for VM relocation for case of WMD attack:

$$Y_{b}^{\gamma} = 0, \forall b \in V_{\gamma}, \quad Y_{b}^{\gamma} \ge \frac{1}{M} \sum_{u \in V_{\lambda}} Z_{u,b}^{\gamma}, Y_{b}^{\gamma} \le \sum_{u \in V_{\gamma}} Z_{u,b}^{\gamma}, \forall b \in (\hat{V} - V_{\gamma}), \gamma \in \Gamma, u \in E \tag{53}$$

Constraint which maps the backup virtual links from any working VM to the backup VM, preparing for relocation in case of a WMD attack:

$$K_{u,b}^{\gamma,e} \le Z_{v,b}^{\gamma}, K_{u,b}^{\gamma,e} \le K_{v,u}^{\gamma,e}, K_{u,b}^{\gamma,e} \ge Z_{v,b}^{y} + K_{v,u}^{\gamma,e} - 1, K_{u,b}^{\gamma,e} = Z_{v,b}^{y} \forall v, u \in V_{\gamma}, b \in (\hat{V} - V_{\gamma}), \gamma \in \Gamma \tag{54}$$

Constraint which limits the number of virtual backup nodes and the capacity of datacenters to relocate VMs in case of a WMD attack:

$$\sum_{b \in (\hat{V} - V_{\gamma})} Y_{u,b}^{\gamma} \ge 2, \sum_{b \in (\hat{V} - V_{\gamma})} Y_{b}^{\gamma} \le \left|V_{\gamma}\right|, \forall \gamma \in \Gamma, \quad P_{free}^{b} - \sum_{u \in V_{\gamma}} P_{u}^{\gamma} Y_{b}^{\gamma} \ge 0 \tag{55}$$

Illustrative example

We tested our approach on a 24-node US mesh WDM optical network with 32 wavelengths per link. Potential WMD attack target regions are shown in Fig. 32(a). The probabilities of attack and damage are based on the cities' populations and importance [4]. We consider five full-mesh cloud networks (CNs), each consisting of four virtual nodes (i.e., VMs) distributed over 16 datacenters (Fig. 32(b)). We assume that each virtual link requires a full lightpath, and each datacenter has enough capacity.

We tested eight approaches: four minimizing resources (Min-Res) and four minimizing risk (Min-Risk). Some of them use a disaster survivable mapping (DS), disaster and post-disaster survivable constraints (D-PDS), and VM backup location (VBL) with number of backup location: one (1L) or two (2L). Min-Res-DS-1L indicates minimization of resources, with 1 VM backup location which we call RESA-1L. The list of approaches is presented in Table 1 including our approaches.

Our examples are evaluated using risk and penalty, disaster and post-disaster survivability, and resource usage analysis. The risk of CN disconnection is evaluated using the first part of Eq. (46). The penalty for capacity loss is the total capacity that can be lost due to a WMD attack. The second analysis is the evaluation of the probability of CN disconnection (PoD). The PoD is calculated by an algorithm called cloud-network resiliency test algorithm (CNRT) which tests the vulnerability of the CN to all possible combinations of disaster and post-disaster failures. CNRT gets the mapping of each CN and simulates disaster damage over the physical infrastructure based on four given failure scenarios – DF: Any single WMD attack occurs without post attack; DSLF: Additional one physical link is affected after the WMD attack; DDLF: Additional two physical links fail after the WMD attack; DFDF: After the first WMD attack, a second WMD attack occurs. Then, the algorithm tests the connectivity of every VM and counts the number of possible failure scenarios caused by a disaster in which the CN is disconnected. With these numbers, CNRT obtains one PoD for each CN and type of failure using Eq. (56).

$$PoD = \frac{Total\ Number\ of\ CN\ Disconnection}{Total\ Number\ of\ Possible\ Failures} \tag{56}$$

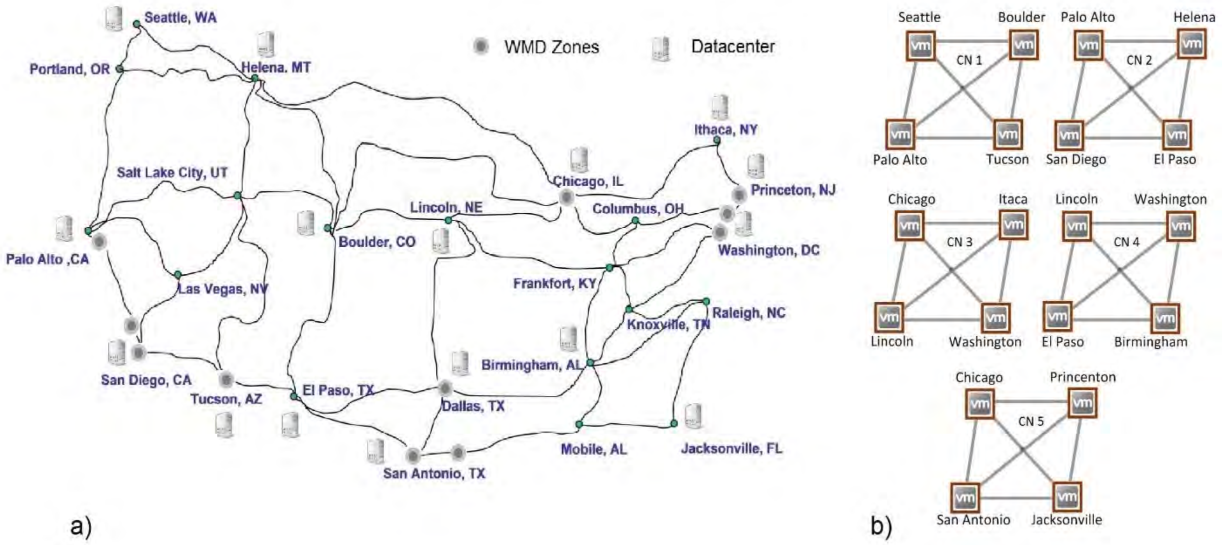


Figure 32. (a) Physical network with WMD zones and datacenters; and b) CNs used in the study.

Name	Approach	PBS	VBL	Cuts
RESA	Min-Res			Basic
RISKA	Min-Risk-DS			Basic
RESA-1L	Min-Res-DS-1L		1L	Basic
RISKA-1L	Min-Risk-DS-1L		1L	Basic
RESA-PDS	Min-Res-D-PDS	X	1L	Extended
RISKA-PDS	Min-Risk-D-PDS	X	1L	Extended
RESA-2L	Min-Res-D-PDS-2L	X	2L	Extended
RISKA-2L	Min-Risk-D-PDS-2L	X	2L	Extended

Table 1. Tested approaches.

To study the risk and penalty, we use the mapping of the five CNs presented in Fig. 32(b). However, we select CN 3 (i.e., CN more affected by DZs) to study the WMD attack and post-attack scenarios (Fig. 33).

(i) RISKA approach reduces the risk of CN disconnection and penalty by 2.75% to 3.77% (i.e., low reduction of risk)

(ii) By adding the VM backup location (VBL), RISKA-1L approach reduces the risk of CN disconnection and penalty up to 88% in most cases. However, VBL works better with RISKA.

(iii) PDS function slightly increases the risk because the extended cuts force virtual links to be mapped in longer lightpaths. However, PDS increases survivability.

(iv) The combination of PDS and VBL with two VM backup locations per CN obtains better survivability.

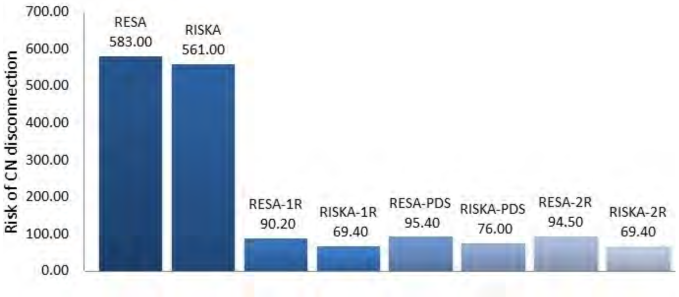
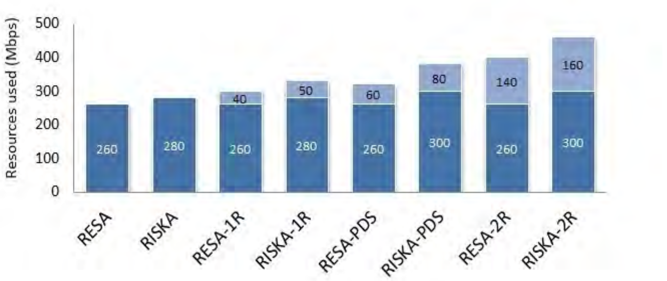


Figure 33. Risk of CN disconnection.

Approach	DF	DSLF	DDLF	DFDF
RESA	0.18	0.38	0.42	0.35
RISKA	0.09	0.38	0.41	0.22
RESA-1L	0	0.29	0.35	0.04
RISKA-1L	0	0.2	0.24	0.02
RESA-PDS	0	0.14	0.17	0
RISKA-PDS	0	0	0.15	0
RESA-2L	0	0	0.17	0
RISKA-2L	0	0	0.15	0

a)



b)

Figure 34. (a) Probability of disconnection (PoD). (b) Resources used to provide resiliency of CN 3.

In terms of probability of disconnection (PoD) of CN 3 (Fig. 34(a)), we observe the following:

(i) DF: CN with VBL will completely survive any failure as any VM can be relocated, i.e., PoD = 0.

(ii) DSLF: RISKA approach reduces PoD up to 22% compared to RESA approach. And, RISKA-1L (i.e., with VBL) increases the survivability by 37% to 100% compared to RESKA-based approaches. PDS constraint increases the survivability to 100% independent of the number of VM backup locations.

(iii) DDLF: When VBL is used, the reduction of PoD is high (between 24% and 64%).

(iv) DFDF: VBL reduces the PoD remarkably by 78% to 100%.

In terms of resources (Fig. 34(b)), we observe that:

(i) RISKA-based approaches require additional resources by 7.8% to 16% to reduce the risk and penalty and PoD. RISKA with VBL constraints increases resource usage by 16% to 37% for one VM backup location (RISKA-1L) to provide risk and penalty reduction by 85% to 87%, and a reduction of the PoD by 24% to 100% (i.e., increasing the survivability by 24% to 100%).

(ii) PDS constraint with RISKA and VM backup location (RISKA-PDS) increase the resources by 23% to 38%. However, the risk and penalty are reduced and survivability is increased up to 100%.

(iii) Two VM backup locations require more resources, but increase the survivability.

In summary, we studied the disaster and post-disaster (e.g., WMD attack and post-attack correlated and cascading failures) survivable cloud network (CN) mapping problem. We proposed a CN mapping approach Min-Risk-D-PDS using (i) VM backup location for each CN (VBL) and (ii) post-disaster survivability constraint (PDS), which offer an economically-sustainable disaster and post-disaster survivable CN mapping approach. We formulated the Min-Risk-D-PDS as an integer linear program. We compared our approach with seven different approaches characterized by different combinations of VBL and PDS constraints with risk and resources minimization as objective function.

Results on a case study formed by five CNs mapped over a US network in a WMD attack case showed that Min-Risk-D-PDS (RISKA-PDS) reduces the risk of CN disconnections and penalty for capacity loss by 85% to 90%. As a consequence, our approach increases the CN survivability by 60% and 100% against three kind of post-disaster failures with the cost of 23% to 50% of additional resources usage. Our illustrative examples confirm the importance of VM backup location and post-disaster survivability constraints for CN survivability against any disaster and post-disaster correlated, cascading failures that may occur in the network.

Protection in Cloud with Manycasting

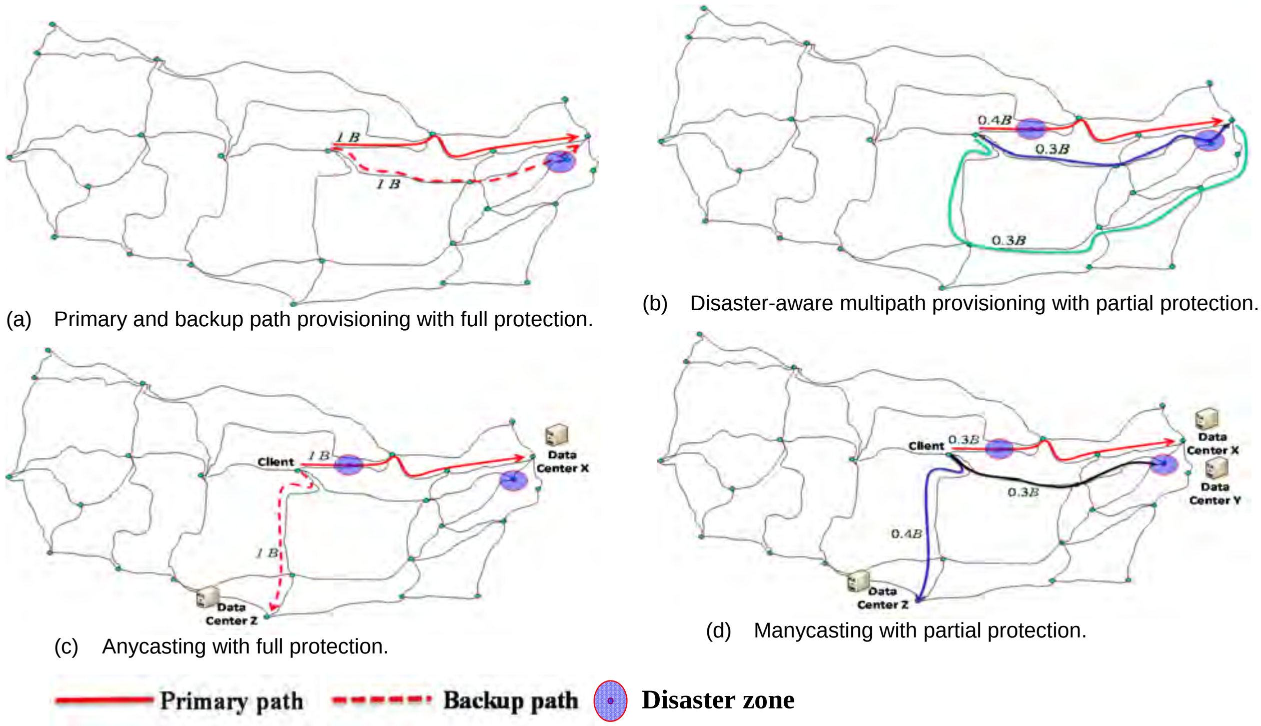
The rapid growth of broadband communications has led to many new web applications such as online interactive maps, social networks (even for the military), video streaming, cloud computing, and CDN (Content Distribution Network) services, most of which are provided by cloud networks which are composed of several datacenters. Cloud-based applications are reshaping the network landscape by pushing the traditional hierarchical and connectivity-oriented Internet towards a more flat and service-oriented infrastructure as promoted in Web 2.0. Service requirements of bandwidth-hungry cloud services make disaster survivability even more crucial as the data loss caused by large-scale correlated, cascading failures can be very high. To alleviate their impact, new measures should be taken as emerging cloud services make different service paradigms possible since one-to-one connectivity is not required anymore. Hence, we have investigated how to increase resiliency by exploiting cloud service paradigm.

Contents or services in cloud systems can be replicated across multiple datacenters located at different nodes from which user demands can be served. As opposed to the traffic that explicitly specifies both source and destination (unicast traffic), cloud networks make new service paradigms feasible which include anycasting, i.e., providing service to cloud users from any of the datacenters that host the service, and manycasting, i.e., providing service from a subset of the datacenters. Our proposed novel WMD-aware service-provisioning scheme multiplexes service over multiple paths to multiple servers/datacenters with manycasting, and the main contribution is to assess the additional resiliency manycasting offers by allowing serving a cloud user from an intelligently-selected subset of datacenters that have the requested content/service.

Figure 35 summarizes the evolution of WMD protection exploiting manycasting. The traditional way is to provide an SRG-disjoint backup path to the connection for WMD protection as in Fig. 35(a). Since providing 100% protection against WMD attacks would be costly as disasters are difficult to predict and statistically rare, some works have showed that multipath provisioning outperforms single-path provisioning in terms of resource usage. As shown in Fig. 35(b), multipath provisioning offers degraded service in case of a failure (partial protection) for throughput-tolerant services which is the case for many cloud applications. These services can continue to operate under reduced capacity (e.g., bandwidth), which results in a lower service/perceived quality (e.g., for file delivery, transfer may take longer; for video conferencing, video frames may have lower quality/resolution with reduced capacity, etc.)

The shifting paradigm towards cloud computing is creating new opportunities for optimizing disaster-aware network design. New service models are enabled such as anycasting, i.e., providing service from any of the datacenters that hosts the requested service, and manycasting, i.e., providing service from a subset of the datacenters. Resilience against destination node failures is very crucial due to cloud services and datacenters hosting content. In our previous work in [11], backup-path protection using anycasting was proposed, as shown in Fig. 35(c). This model places datacenters and routes services under the anycast model, such that fast cloud service protection is considered in case of single-link and single-node failure events. Our manycasting solution in Fig. 35(d) combines the benefits of multipath and anycast provisioning schemes during the delivery of cloud services. Our proposed manycasting protection offers (1) degraded service in case of large-scale failures due to disasters as in multipath provisioning and (2) resilience to datacenter node failures as in using a backup-datacenter approach.

We propose to provision multiple paths, which are the least risky ones, to multiple datacenters, which we call Multipath to Multiple Destinations (MMD). We define risk using a probabilistic model i.e., we consider disaster occurrence probabilities as well as the probability of the network resources being affected by a disaster. The main benefits and properties of the proposed scheme are as follows. First, it provides protection against secondary failures (e.g., aftershocks) along with single-link and disaster failures by provisioning multiple SRG-disjoint paths. When a connection loses one path after a disaster, it will still have multiple SRG-disjoint paths which protect it against post-disaster failures. Second, it is resilient against destination node failures (since we have multiple destination nodes, the service will not be disrupted if one of them fails due to a disaster). Third, this scheme can ensure degraded service (vs. no service at all) after a failure without using extra resources. We compare our scheme with two benchmark schemes, namely Multipath to Single Destination (MSD) and Backup path to Backup Destination (BBD). In MSD, multiple paths which use a single datacenter are provisioned for a request, i.e., multipath provisioning. In BBD, one primary and one backup path are provisioned for a connection in which both the paths and their destination nodes are SRG disjoint.



(a) Primary and backup path provisioning with full protection.

(b) Disaster-aware multipath provisioning with partial protection.

(c) Anycasting with full protection.

(d) Manycasting with partial protection.

Figure 35. Disaster-aware provisioning schemes: (a) Full protection via backup path (Partial protection can be realized by lowering bandwidth of backup path), (b) Partial protection via multipath routing, (c) Full protection via anycasting (partial protection also possible), and (d) Partial protection via manycasting.

Mathematical Formulation

We consider a WDM optical backbone network where datacenters can be placed at a selected subset of network nodes. Contents are not fully replicated (i.e., a content exists in only a subset of the datacenters) since it may create more background traffic between datacenters due to synchronization.

Disaster-aware manycast routing can be stated as follows: *Given the network topology with nodes and fiber links interconnecting them, including source sites (where requests for cloud services originate), candidate destinations (i.e., the cloud servers), and disaster risk map of the network, find multipaths to multiple datacenters for each of the requests such that the risk (i.e., expected bandwidth loss) of the network during a disaster event is minimized.*

The mathematical formulation of multipath provisioning to multiple datacenters is formally described below, and it turns out to be an Integer Linear Program (ILP). Datacenter locations and content placement are given as input. At most one path will be established to a datacenter for a connection. We consider k-shortest paths from each node to each datacenter as input. The primary objective is to find a set of paths for all connection requests that minimizes the expected bandwidth loss in the event of disasters and the secondary objective is to minimize network resource usage.

Input Parameters:

$G(V,E)$: Network topology where V is the set of nodes and E is the set of directed links.

$P = \{p \mid p = <s_p, d_p, L_p, E_p>\}$: Set of k-shortest paths where s_p, d_p, and L_p are source, destination, and length of the path in km, respectively. E_p is the set of links on path p.

$D = \{d \mid d = <v_d, C_d>\}$: Set of datacenters where v_d is the node where datacenter d is located and C_d is the set contents datacenter *d hosts*.

$T = \{t \mid t = <s_t, c_t, B_t, \mu_t, n_t, P_t>\}$: Set of connection requests where s_t is the source, B_t is the bandwidth request, μ_t is the partial protection ratio (0.5 for 50% protection and 1.0 for full protection), n_t is the minimum degraded service ratio (for 30% minimum protection, this value is 0.30), c_t is the requested content of connection t, and $P_t \subset P$ is the set of possible paths to use for connection t where $s_t = s_p$.

$Y = \{y \mid y = <E_y, \rho_y>\}$: Set of WMD zones where E_y is the set of links that is member of WMD region y and ρ_y is the probability that WMD attack y causes a failure.

K: Maximum number of paths allowed for a connection.

$U_p^y \in \{0,1\}$: Equal to 1 if path p goes through SRG y.

$A_p^e \in \{0,1\}$: Equal to 1 if link e $\in E_p$.

$H_{c_t}^{d_p} \in \{0,1\}$: Equal to 1 if datacenter d has content c.

π_e: Capacity of link e (different for each link).

DDC: Differential delay constraint.

Integer Variables:

α_p^t: Bandwidth used on path p for connection t.

X_t: Number of paths for connection t.

$Z_p^t \in \{0,1\}$: 1 if path p is used for connection t.

Objective: Minimize

$$\sum_{y\in Y}(\sum_{t\in T}(\sum_{p\in P_t}\alpha_p^t\ .U_p^y).\rho_y + \in.\sum_{t\in T}\sum_{p\in P_t}\alpha_p^t + Z_p^t.L_p$$

The first term minimizes the expected bandwidth loss (risk) in case of WMD attacks. The second term ensures that shortest paths are selected and minimum possible bandwidth is provisioned.

Constraints:

$$\sum_{p\in P_t} \alpha_p^t \geq B_t \, \forall t \in T \tag{57}$$

$$\sum_{p\in P_t} \sum_{t\in T} \alpha_p^t . A_p^e \; \leq \; \pi_e \, \forall e \in E \tag{58}$$

$$\sum_{p\in P_t} Z_p^t = X_t \tag{59}$$

$$X_t \leq K, X_t \geq 2 \quad \forall t \in T \tag{60}$$

$$Z_p^t \leq \alpha_p^t \forall t \in T \text{ , } \forall p \in P_t \tag{61}$$

$$Z_p^t \geq \alpha_p^t / M \qquad \forall t \in T \;\; \forall p \in P_t \tag{62}$$

$$Z_p^t \leq \frac{H_{c_t}^{d_p} + 1}{2} \qquad \forall t \in T \, \forall p \in P_t \tag{63}$$

$$\sum_{p\in P_d} Z_p^t \leq 1 \qquad \forall t \in T \forall d \in D \tag{64}$$

$$\sum_{p\in P_t} Z_p^t \, . U_p^y \leq \; 1 \;\; \forall t \in T, \, \forall y \in Y \tag{65}$$

$$\sum_{p\in P_t} Z_p^t \, . A_p^e \leq \, 1 \forall t \in T \text{ , } \forall e \in E \tag{66}$$

$$|(L_p - L_q).Z_p^t . \; Z_q^t| \leq DDC \;\; \forall p, q \in P_t \;\; (p \neq q), \forall t \in T \tag{67}$$

$$\sum_{p\in P_t} \alpha_p^t - \; \sum_{q\in P_t} \alpha_q^t \, . U_q^y \geq \mu_t . B_t, \, \forall t \in T, \forall q \in P_t, \, \forall y \in Y \tag{68}$$

$$\alpha_p^t \geq Z_p^t . nt . Bt \qquad \forall \mathrm{t} \in \mathrm{T}, \forall \mathrm{p} \in \mathrm{P_t} \tag{69}$$

Equation (57) enforces that the total reserved bandwidth is at least equal to the requested bandwidth. Eqn. (58) ensures that the total bandwidth usage in a link does not exceed the link capacity. Eqns. (59) and (60) constrain the number of multipaths per connection. Eqns. (61) and (62) define Z_p^t where M is a large number. Eqn. (63) ensures that only paths destined to datacenters with the requested content are selected for a connection. Eqn. (64) ensures that at most one path is provisioned to a specific datacenter for a connection. Eqns. (65) and (66) ensure that every path used for a connection is SRG-disjoint and link-disjoint, respectively. As the number of multipaths increases, the survivability against post-disaster failures also increases. Possible multipath number is dependent on the availability of the network resources, content replication amount, as well as the DDC. Eqn. (67) ensures that the differential path distances of a connection are selected to fulfill DDC (whose practical value is 8 ms). Eqn. (68) ensures that, in case of single-link or disaster failures, requested level of degraded service is provided. When the requested degraded service ratio is 1, full protection with overprovisioning is provided. Eqn. (69) gives the minimum level of degraded service required in case only one path survives (e.g., if a connection's multipath number is 4, it will take minimum degraded service after any three simultaneous failures).

```
Input: G( V, E ), T= { t | t= <$s_t$, $b_t$ , $\mu_t$ ,$c_t$>}

procedure initialize( void ):
//Build a bipartite graph
set N,D to empty set
foreach (v in V)
    if (v is a datacenter)
        D.add(v)
    else N.add(v) endif
endforeach
foreach (n in N)
    foreach (d in D)
    $K_{nd}$ = kshortest(n, d, k)
    sortPathsbyRisk($K_{nd}$, +$C_{n,d}^{k}$ ∀ k)
    $Z_{nd}$ = select the least risky m paths between nodes {n,d}
    endforeach
endforeach
foreach (t in T)
    provision(t) endforeach

procedure provision(t):
Set D' to an empty set.
foreach (d in D)
    if ($C_d$ contains $c_t$)
    D'.add (d) endif
endforeach
set solution set S to empty set.
set numMaxCon = min(size(D'), 4)
for i = 2 to numMaxCon
  set D" to Combination(D',i)
//D" is vector of datacenter set.
  set sol to findSolution(t,D")
  S.add (sol)
endfor
sortCostofSolutionSet(S)
return S[0]

procedure findSolution (t, D")
set S' to empty set. //sol. set
foreach ($D_s$ in D")
  foreach possible combinations of m least risky paths
      Let $P_t$ be the selected paths to datacenters in $D_s$ among all $Z_{nd}$ where n is $s_t$ and d is in
        $D_s$.
      if(assignCapacity(t,$P_t$)*)
      set Solution S to $P_t$ with   capacities assigned.
      set costOfS by Eq.(14)
      S'.add(S)
  end_foreach
end_foreach
sortCostofSolutionSet(S')
return S'[0]
```

Figure 36. Pseudo code for proposed heuristic.

+ $C_{n,d}^{k}$ is the risk of the kth path between *n* and *d* shown in Eq. (71).

* We use a modified version of capacity assignment algorithm in [17]. It assigns bandwidths to the paths by satisfying degraded service constraint.

Heuristic

Solving the ILP is not an easy task for large-scale problem instances. To make the design more scalable for large problem instances, we design a heuristic. We use this heuristic for static traffic, where all requests are known beforehand, though it can also be applied for dynamic traffic, where requests arrive and are processed one-by-one. We provide the pseudo code of the proposed heuristic in Fig. 36 to provision a set of connection requests T.

The heuristic starts with initialization steps (lines 1-16). First, we construct a bipartite graph (see Fig. 37) in which the source nodes are on the left set and datacenters on the right. Second, we pre-calculate the k-shortest paths from each possible source to each datacenter in order to make future calculations more efficiently. All procedures starting with "sort", are regular sort operations (e.g., sortPathsbyRisk sorts connections based on their risk level, sortCostofSolutionSet sorts provisioning solutions for the connections request based on the cost of the paths with assigned capacity.) Once the initialization is complete, we switch to the running state where, for each request t, we invoke the provision procedure. The procedure starts with determining the candidate datacenters which can serve the requested content (lines 21-24) and form the set D'. The next step is to determine possible datacenter sets to provision a connection. Since we limit the system with maximum of four paths per connection, all possible two-to four-elements combination of D' is determined. For this purpose, we first construct the set D" which is the vector of all possible set of candidate datacenters using the combination operation. For instance, if i (number of connections) is set to two and if the content is available in three servers (A, B, and C), then, D" is the set of ({A, B}, {A, C}, {B, C}, {A, B, C}). The basic idea is to forward set D" to the findSolution procedure along with the incoming connection request, record the results (lines 27-32), and then sort all results based on the cost function to choose the one with the minimum cost (lines 27-34). (For sake of simplicity, we omit the case where no possible result exists.) In findSolution procedure (lines 36-49), we use a similar iterative method as above and break D" into subproblems, calculate the cost for each possible solution, and then choose the one with the lowest cost again. At this point, the problem boils down to finding the best (lowest-cost) solution for a given set of candidate datacenters (e.g., A, B) which correspond to determining the paths from the source to each datacenter, assigning capacity to each path and calculating the cost. We iterate over all possible path combinations for a datacenter set to which we are trying to provision paths for a connection.

A set of possible paths between a source to a datacenter set is called a solution. We assign capacities to every solution using capacity assignment in [17] while satisfying the connection request's bandwidth and degraded service requirement:

$$C_s = \sum_{y\in Y}\left(\sum_{p\in P_s} \alpha_p^s \,.\, U_p^y\right).\, \rho_y + \in .\sum_{p\in P_s} \alpha_p^s \quad (70)$$

$$C_{n,d}^k = \sum_{y\in Y} U_p^y \cdot \rho_y \quad (71)$$

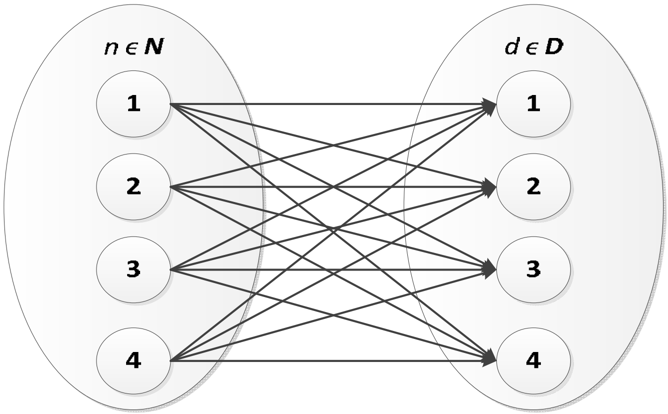


Figure 37. Bipartite graph: Each edge <*n*, *d*> represent k-shortest paths between nodes {*n, d*}, and each path $P_{n,d}^k$ is weighted by its risk of failure found in Eqn. (71).

At the last step, we calculate the cost of the solution using the Eqn. (70) which is the expected bandwidth loss in case of all predefined disasters and the total bandwidth provisioned for the connection. Even if the number of loops seems to be high, given that k may be bounded to a low value and a requested content is available in a limited number of datacenters, the computations are done efficiently, and in a much faster way compared to the ILP.

llustrative Numerical Examples

We present illustrative results by solving the ILP formulations. We compare three WMD-aware provisioning schemes: multipath to single destination (MSD), our proposed multipath to multiple destination (MMD), and backup path to backup destination (BBD) (i.e., SRG-disjoint primary and backup paths to SRG-disjoint datacenters), in terms of resource usage and survivability. All three schemes perform routing in a risk-aware manner (i.e., they select their paths to reduce the expected loss in case of a disaster). In all schemes, all paths are link-disjoint and SRG-disjoint. The ILP formulations for MSD and BBD are derived from our MMD model with minor changes.

We study a 24-node US-wide network (Fig. 38) with 32 wavelength channels and 10 Gbps per channel per link. As disaster scenarios, we consider WMD attacks and 10 most-populated cities and Washington DC as possible targets, as in [7]. WMD zones are shown in circles. We assume that, if a disaster occurs, all nodes/links in the affected zone are down. We consider 5 datacenters as shown in Fig. 38 placed at safe nodes (not in a WMD zone), which minimizes the total distance of each node to its nearest datacenter. For datacenter placement, we use the greedy algorithm for the traditional Center Selection problem. In this study, as ILP is computationally extensive, we consider a small example with 15 contents and 3 geographically-distributed replicas per content. The bandwidth distribution of the connection requests is OC-3 (~150M), OC-12 (~600M), OC-48 (~2.5G), and OC-192 (~10G) which follows the ratios 40: 30: 20: 10 (which is a realistic bandwidth distribution in a practical network).

Content replica number limits the maximum number of multipath to 3. For nodes with degree less than 3, it is infeasible to find 3 link-disjoint paths; so, fewer paths will be provisioned for them. To compare our schemes, we only considered connections that can have 3 or more SRG-disjoint paths.

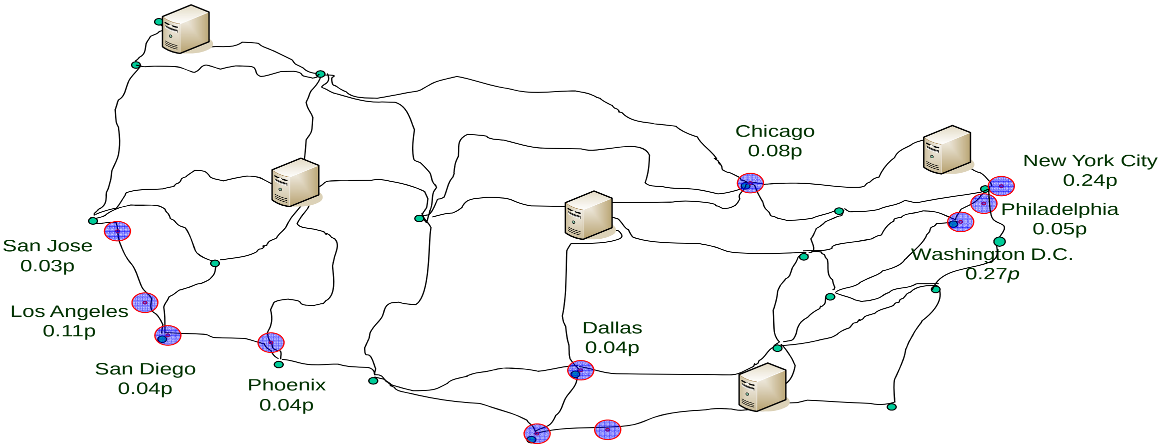
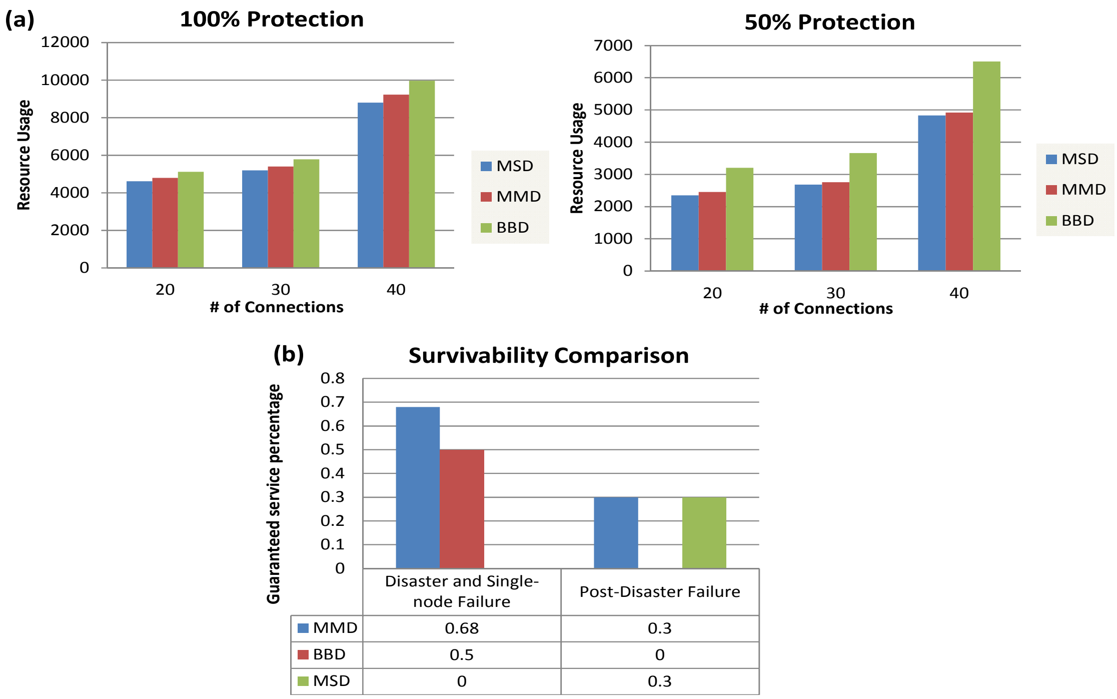


Figure 38. US-wide network with WMD disaster zones shown in circles with attack probabilities, where *p* is the probability of a WMD attack targeting US.



MMD → manycasting MSD → multipath BBD → anycasting

Figure 39. Comparison of the proposed scheme with benchmarks using ILP: (a) Resource usage comparison of different risk-aware provisioning schemes with same protection (100% and 50%) against disaster failures, (b) Guaranteed service amount (worst-case scenario).

Figure 39(a) compares the total resource usage of our scheme with MSD and BBD schemes where all schemes provide same level of protection. Our scheme (MMD) consumes approx. 25% less resources than BBD in case of 50% protection, and 10% less resources in case of full protection while providing same amount of survivability against single-link and single-node (with a WMD attack) failures. However, while BBD may become vulnerable to even single-link failures after a WMD attack, our scheme offers survivability against post-WMD failures as well. Resource consumption of MMD and MSD are very close. In this particular setting, only around 5% more resource is consumed by MMD. In return, it provides protection against node failures, while in MSD, connectivity may be lost. The resource usage directly depends on the number of datacenters and their geo-distribution as our approach tries to provide service from as many datacenters as possible satisfying DDC. In Fig. 39(b), we compare the schemes according to their guaranteed minimum service amount in different failure scenarios (WMD, post-WMD, and node failure) when the resource usage is almost same (but the *guaranteed* service levels they offer differ). Tests are done with 20, 30, and 40 connections; since results are similar, we provide only the results with 30 connections. Resource usage is defined as the total bandwidth consumed on all links in the network. For example, if a connection path with 3 links has 5B capacity, then it uses 15B in total. We set all schemes' resource usage to around three values in terms of optical carrier (OC): 3500, 4500, and 5500, and resource usages of all schemes are within 2% confidence interval. We observe that our scheme (MMD) provides 10-20% more protection than BBD in all cases. MSD and MMD results are similar except in node failures. If datacenters are not located in risk-free nodes, then MSD cannot guarantee any degraded service in case of node failures, because all connection bandwidth can fail if the datacenter it connects is down. To sum up, our scheme offers comprehensive protection for different failure scenarios.

Since the ILP approach is not scalable, we can optimally solve the ILP only in small problem instances. To demonstrate the effectiveness of our heuristic, we compare the heuristic and ILP results. We compare ILP's resource usage with that of the heuristic on the 24-node network for 20, 30, and 40 connections, 5 datacenters as shown in Fig. 38, and 3 replicas per content. Figure 40(a) shows that our heuristic's resource usage has a similar trend with ILP. The small gap-to-optimality of around 20% confirms the superior performance of the heuristic. Figure 40(b) shows that, when the same amount of resources is provided for both ILP and heuristic, their survivability characteristics are comparable.

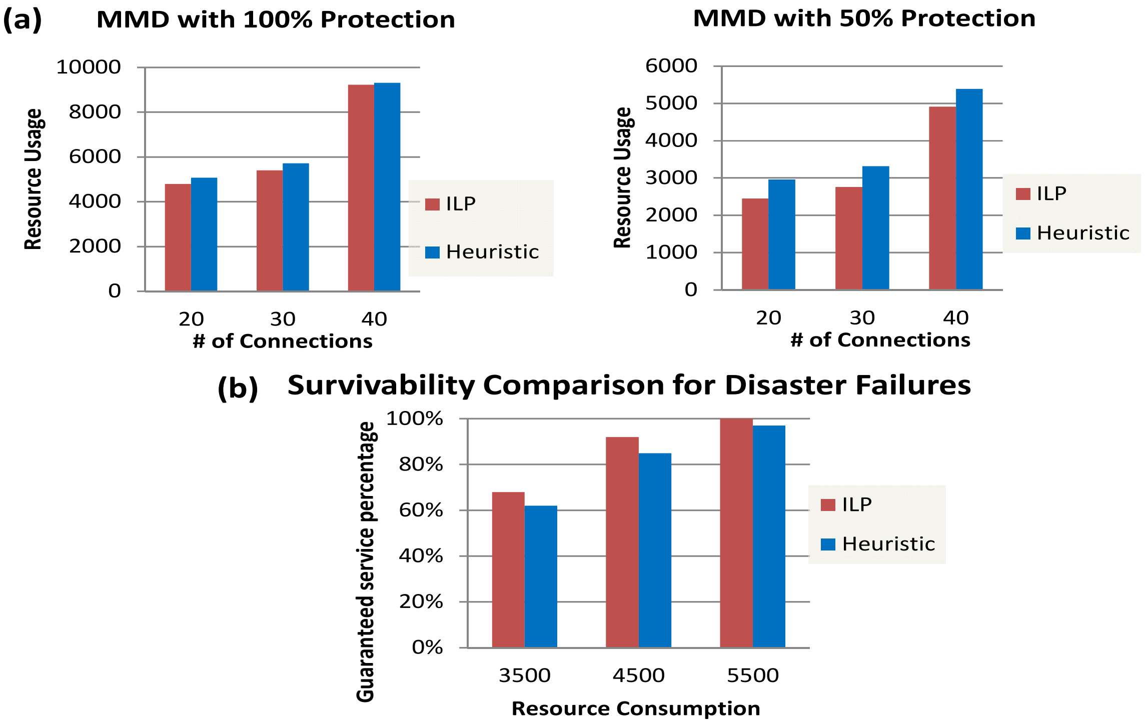


Figure 40. ILP vs. heuristic comparison: (a) Resource usage comparison of ILP and heuristic approaches for MMD scheme, (b) Guaranteed service amount comparison between ILP and heuristic when they consume same amount of bandwidth.

3. What opportunities for training and professional development has the project provided?

The project supported the research of the following students under the supervision of the PI (Mukherjee):

1) Dr. Ferhat Dikbiyik was a PhD student in the PI's lab, funded between 2010 and 2013 under this grant, and worked on characterization of WMD stressors, WMD-aware reprovisioning, and multi-path provisioning.
2) Dr. M. Farhan Habib, was a PhD student in the PI's lab, funded between 2010 and 2014 under this grant who worked primarily on content connectivity in cloud and protection of virtual networks against WMD attacks and natural disasters.
3) Ms. Saadet Sedef Savas, PhD Candidate, who has been working on degraded service and multipath based protection of content in cloud.
4) Mr. Carlos Colman Meixner, PhD Candidate, who has been working on protection of virtual networks.
5) Ms. Sifat Ferdousi, PhD Candidate, who has been primarily working on WMD-resilient content management in the cloud.
6) Mr. Abhishek Roy, PhD Candidate, who has been working on the problem of resource crunch due to WMD disasters.

<u>This project supported the following visiting researchers:</u>

1) Dr. Massimo Tornatore from Politecnico di Milano, Italy, was a Visiting Postdoctoral Researcher in the PI's lab, and is current adjunct professor at UC Davis. He helped us to develop techniques for disaster-resilient data replication and formulate the problem.

2) Dr. Marc De Leenheer from Ghent University, Belgium, was a Visiting Postdoctoral Researcher in the PI's lab. He helped us understand the problem on data replication with protection and how to formulate the problem. He also helped us with formulation of WMD-aware provisioning problem.

Other collaborators and contacts involved in this research:

1) Dr. Ning-hai Bao from Chongqing University of Posts and Telecommunications, China, who worked on some aspects of disaster protection.
2) Dr. Abu Sayeem Reaz: Park, Vaughan & Fleming LLP, Davis, CA, and former PhD student in the PI's lab. Dr. Reaz helped with the formulation of WMD-aware provisioning problem.

Every year, the trainees attended the annual Optical Fiber Communication Conference (OFC) to increase their knowledge on the relevant topics.

4. Impact

To better quantify the improvements of this research over the current state of the art, we have summarized some of the contributions below:

We proposed risk-aware provisioning methods which choose less-risky regions for connections during provisioning to reduce the loss in case of attacks. We evaluated our approaches in terms of risk and penalty reduction and increase in (consumed) resources compared to an approach which minimizes resource consumption. Risk reduction is between 25% and 35% with increase in resources around 25%. Risk-aware provisioning requires more resources than risk-unaware provisioning. However, excess capacity (EC) available in the networks (considering low utilization of the current networks) can be exploited to provide better protection against WMD disruptions (e.g., avoid risky zones, whenever possible, by using the EC of the links traversing safe zones).

After an attack, some of the traffic will be disrupted. We proposed a disaster-aware reprovisioning method which considers correlated cascading failures and sequential attacks which might disrupt more network components and connections. Our numerical examples showed a 26% reduction in penalty (a value to understand the loss of network operator in case of an attack) for an illustrated WMD attack on Washington DC.

To remedy the dramatic effects of the WMD attacks on the available network resources in the network, we proposed to exploit service requirements of the connections, and provide degraded service (refers to the reduced amount of resource allocation for a service vs. its normal operation requirement), to some services which can still achieve lower but acceptable quality. At high network loads, this approach provides as much as 10% reduction in the amount of blocked bandwidth (a.k.a, connections) in return of 3% decrease in the average provided bandwidth.

To ensure that cloud networks get minimally affected (disconnected) by failures in the physical network caused by large-scale terrorist attacks, we proposed a disaster-survivable cloud-network mapping technique, which yields risk and penalty reduction by 85% to 87%.

To decrease the risk of content loss due to unavailability or unreachability in the cloud networks due to WMD attacks, we proposed an approach to determine the number of the replicas required and their locations. Our results show that the WMD-aware approach gives significant improvement in terms of risk reduction (up to 45%), for a given example. We also see that our approach performs better with increasing number of contents.

To better prepare the network given the increasing risk of an upcoming attack, we proposed a rapid-data evacuation method to save as much data as possible in cloud networks. Compared to evacuating the data to the closest datacenter, our intelligent rapid-evacuation approach evacuates data much faster, and takes 63% less time to evacuate same amount of data which is a significant time savings in case of attacks.

As the effect of WMD attacks on communication networks is directly related to their effect on the physical infrastructure, determining the exact locations of the components of a network, and combating the attacks given this information are very significant. We have constructed a model of a sample US nation-wide backbone optical network with physical locations of all vertices and links by matching US-wide topology with US transportation map, with the assumption of “communication lines are close to highways and/or railroads”.

A. Impact on other disciplines

Our research group enjoys support from industry such as Teknovus/Broadcom, NTT, Nokia, Fujitsu, ETRI Korea, NEC, Siemens, HP Labs, and Huawei. Also, as broader influence of his federally-funded research, the PI serves or has recently served as a Member of the Board of Directors or on the Technical Advisory Board of several companies: IPLocks (acquired by Fortinet), Teknovus (acquired by Broadcom), Intelligent Fiber Optic Systems (IFOS), and LookAhead Decisions Inc. (LDI). Nokia Siemens Networks (NSN), a global network equipment vendor has supported our research on the related problem of reliable architectures for Carrier Ethernet networks. ETRI Korea, Huawei, and HP Labs are collaborating with us on defining reliable architectures for integrated wireless-wireline access networks.

B. Impact on physical, institutional, and information resources that form infrastructure

The PI’s group frequently publishes tutorially-oriented articles in widely-read journals and magazines such as IEEE Communications Surveys and Tutorials, IEEE Network, IEEE Communications Magazine, etc. PI Mukherjee serves as the Series Editor of Springer’s “Optical Networks Book Series” which publishes a variety of books on optical networking, including Robust Network Design, under his supervision and guidance.

C. Impact on technology transfer

The techniques developed for WMD-aware reprovisioning are also applicable for any disaster-aware provisioning on telecom networks. Major recent disasters (e.g., 2012 Hurricane Sandy and 2011 Japan earthquake) has been leading telecom companies to seek for more robust solutions, and we believe our approach is one of the solutions. The disaster-resilient data replication technique is suitable for datacenter networks where both path protection and content protection should be considered.

D. Impact on society beyond science and technology

Alumni from our lab – who have been primarily trained by federal research support over the past 25+ years – have been gainfully employed in various companies and universities. Nearly a dozen of them have been women. Three of the PI’s PhD alums – Dr. Shun Yao, Dr. Laxman Sahasrabuddhe, and Dr. Abu Ahmed Sayeem Reaz – chose to pursue law. The first two have finished law school and are practicing lawyers. Three other alums – Dr. Dhritiman Banerjee, Dr. Canhui (Sam) Ou, and Dr. (Ms.) Xiaoling Qiu – are working in the finance sector where they are utilizing their technology/telecom background. Another alum, Dr. Glen Kramer, is Chief Scientist of Teknovus, a startup company based in Petaluma CA, which was recently acquired by Broadcom, and he plays a very important role in the global FTTH business, being the chairperson of the IEEE standards effort for 10G EPON. Ms. Guoying Zhang, Visiting Scientist in our lab for the past 12 months, just returned to her job at the China Academy of Telecom Research, and she will continue to play a significant role in influencing worldwide telecom standards through the ITU and the IEEE. Thus, federal support of our research program over the past two decades is having strong impact in disciplines beyond science and engineering.

5. How have results been disseminated to communities of interest?

The PI's group frequently publishes tutorial articles in widely-read journals/magazines such as IEEE Communications Surveys and Tutorials, IEEE Network, IEEE Communications Magazine, etc. The PI serves as the Series Editor of Springer's "Optical Networks Book Series" which publishes a variety of books on optical networking, including Robust Network Design, under his supervision and guidance.

A. Keynote, invited, and other talks

1. B. Mukherjee, "Extremely large data transfer using the cloud: how it may affect the DoD and DTRA," **DTRA**, Washington DC, June 19, 2014.
2. B. Mukherjee, "Network Convergence in the Future Internet," **Keynote Talk** presented at 12th International Conference on High Performance Switching and Routing (HPSR), Cartagena, Spain; July 4, 2011.
3. B. Mukherjee, "Some Opaque Problems in Transparent Optical Networks," **Keynote Talk** presented at International Conference on Transparent Optical Networks (ICTON), Stockholm, Sweden; June 27, 2011.
4. B. Mukherjee, "Panorama of Optical Network Survivability," **Keynote Talk** presented at Optical Network Design and Modeling (ONDM) Conference, Brest, France; April 18, 2013.
5. B. Mukherjee, "Network Adaptability from Disaster Disruptions and Cascading Failures," Invited Talk presented at **FCC Workshop on Network Resiliency**, New York; Feb. 6, 2013.
6. B. Mukherjee, "Network Adaptability from Disaster Disruptions and Cascading Failures," **Tutorial Talk** at Asia Communications and Photonics (ACP) Conference, Beijing, China; Nov. 15, 2013.
7. B. Mukherjee, B. Mukherjee, "Disaster preparedness for network and information infrastructures," **Invited Talk**, *Optical Network Design and Modeling Conference*, Kista, Sweden, May 22, 2014.
8. B. Mukherjee, "Disaster preparedness for network and information infrastructures," **Guest Professorship Lecture**, *Beijing University of Posts and Telecommunications (BUPT),* Beijing, China, June 26, 2014.

B. Journal papers

1. S. Huang, M. Xia, C. Martel, and B. Mukherjee, "A multistate multipath provisioning scheme for differentiated failures in telecom mesh networks," *IEEE/OSA Journal of Lightwave Technolgy*, vol. 28, no. 11, pp. 1585-1596, 2010.
2. F. Dikbiyik, L. Sahasrabuddhe, M. Tornatore, and B. Mukherjee, "Exploiting excess capacity to improve robustness of WDM mesh networks," *IEEE/ACM Transactions on Networking*, vol. 20, no. 1, pp. 114-124, February 2012.
3. M. F. Habib, M. Tornatore, M. De Leenher, F. Dikbiyik, and B. Mukherjee, "Design of disaster-resilient optical datacenter networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 30, no. 16, pp. 2563-2573, 2012.
4. F. Dikbiyik, M. Tornatore, and B. Mukherjee, "Adaptive time- and location-aware routing in telecom mesh networks," *IET Networks*, vol. 2, no. 1, pp. 19-29, Feb. 2013.
5. M. F. Habib, M. Tornatore, F. Dikbiyik, and B. Mukherjee, "Disaster survivability in optical communication," *Computer Communications*, vol. 36, no. 6, pp. 630–44, 2013.
6. B. Mukherjee, M. F. Habib, and F. Dikbiyik, "Network adaptability from disaster disruptions and cascading failures," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 230-238, May 2014.
7. F. Dikbiyik, M. Tornatore, and B. Mukherjee, "Exploiting Excess Capacity for Survivable Traffic Grooming in Optical Backbone Networks," *Journal of Optical Communications and Networking,* vol. 6, no. 2, pp. 127-137, Feb. 2014.
8. C. Colman-Meixner, F. Dikbiyik, M. F. Habib, M. Tornatore, Chen-Nee Chuah, and B. Mukherjee, "Disaster-survivable cloud-network mapping," *Photonic Network Communications*, vol. 27, no. 3, pp. 141-153, June 2014.
9. F. Dikbiyik, M. Tornatore, and B. Mukherjee, "Minimizing the risk from disaster failures in optical backbone networks," *IEEE/OSA Journal of Lightwave Technology,* vol. 32, no. 18, Sept. 2014.

10. S. S. Savas, F. Dikbiyik, M. F. Habib, and B. Mukherjee, “Network adaptability to disaster disruptions by exploiting degraded-service tolerance,” *IEEE Communications Magazine*, vol. 52, no.12, Dec. 2014.
11. S. Savas, F. Dikbiyik, M. F. Habib, and B. Mukherjee, “Disaster-aware service provisioning with manycasting in cloud networks,” *Photonic Network Commun.*, vol. 28, no. 2, pp. 123-134, 2014.
12. S. Ferdousi, F. Dikbiyik, M. F. Habib, M. Tornatore, and B. Mukherjee, "Disaster-Aware Datacenter Placement and Dynamic Content Management in Cloud Networks," *IEEE/OSA Journal of Optical Communications and Networking*, (under review).

C. Conference papers

1. F. Dikbiyik, M. Tornatore, L. Sahasrabuddhe, and B. Mukherjee, “Management of Excess Capacity for Path-Oriented Differentiated Services Optical Networks,” *Proc., IEEE Photonics in Switching Conference* 2010, Monterey, CA, Aug. 2010.
2. M. Farhan Habib, M. Tornatore, M. De Leenheer, F. Dikbiyik, and B. Mukherjee, “A Disaster-Resilient Multi-Content Optical Datacenter Network Architecture,” *International Conference on Transparent Optical Networks*, Stockholm, Sweden, June 2011.
3. M. Tornatore, F. Dikbiyik, and B. Mukherjee, “(3W-)Availability-Aware Routing in Optical WDM Networks: When, Where and at What Time,” *International Conference on Transparent Optical Networks*, Stockholm, Sweden, June 2011.
4. F. Dikbiyik, M. Tornatore, and B. Mukherjee, “Exploiting excess capacity for survivable traffic grooming in optical WDM backbone networks,” *IEEE Global Communication Conference*, Houston, TX, Dec. 2011.
5. F. Dikbiyik, M. Tornatore, and B. Mukherjee, “Optimal relocation of excess capacity in optical WDM networks,” *IEEE Global Communication Conference*, Anaheim, CA, Dec. 2012.
6. F. Dikbiyik, M. De Leenheer, A. S. Reaz, and B. Mukherjee, “Minimizing disaster risk in optical telecom networks,” *Optical Fiber Communications Conference*, Los Angeles, CA, March 2012. (Semifinalist for Corning Outstanding Student Paper Award; ranked in top 2 out of over 100 submissions in the Core Networking area.)
7. M. F. Habib, M. Tornatore, and B. Mukherjee, “Fault-tolerant virtual network mapping to provide content connectivity in optical networks,” *Optical Fiber Communications Conference*, Anaheim, CA, March 2013. (Semifinalist for Corning Outstanding Student Paper Award; ranked in top 2 out of over 100 submissions in the Core Networking area.)
8. C. Meixner, F. Dikbiyik, M. Tornatore, C-N. Chuah, and B. Mukherjee, "Disaster-Resilient Virtual-Network Mapping and Adaptation in Optical Networks," *Optical Network Design and Modeling conference*, Brest, France, April 2013.
9. S. Ferdousi, F. Dikbiyik, M. F. Habib, M. Tornatore, and B. Mukherjee, “Disaster-aware dynamic content placement in optical cloud networks,” *Proc. Optical Fiber Communications (OFC) Conference*, March 2014, San Francisco, CA. (Promoted in OFC press release.)
10. S. S. Savas, M. F. Habib, M. Tornatore, and B. Mukherjee, “Exploiting degraded-service tolerance to improve performance of telecom networks,” *Proc., Optical Fiber Communications (OFC) Conference*, March 2014, San Francisco, CA.
11. S. S. Savas, F. Dikbiyik, M. F. Habib, and B. Mukherjee, “Disaster-aware service provisioning by exploiting multipath routing with manycasting in telecom networks,” *Proc. IEEE Advanced Networks and Telecommunication Systems (ANTS) Conference*, Chennai, India, Dec. 2013.
12. S. Ferdousi, F. Dikbiyik, M. F. Habib, and B. Mukherjee, “Disaster-aware data-center and content placement in cloud networks,” *Proc. IEEE Advanced Networks and Telecommunication Systems (ANTS) Conference*, Chennai, India, Dec. 2013.
13. S. Ferdousi, M. F. Habib, M. Tornatore, and B. Mukherjee, “Rapid Data Evacuation for Large-Scale Disasters in Optical Cloud Networks,” *Proc., Optical Fiber Communications (OFC) Conference*, Los Angeles, CA, March 2015.
14. A. Roy, M. Farhan Habib, and Biswanath Mukherjee, “Network Adaptability under Resource Crunch: Availability and Revenue Maximization,” *Proc. IEEE Advanced Networks and Telecommunication Systems (ANTS) Conference*, New Delhi, India, Dec. 2014.

References

[1] Office of Management and Budget, US Government, “A year of change in federal IT,” http://www.whitehouse.gov/blog/2011/12/08/year-change-federal-it.

[2] M. F. Habib, M. Tornatore, F. Dikbiyik, and B. Mukherjee, “Disaster survivability in optical communication networks,” *Computer Communications*, vol. 36, no. 6, March 2013.

[3] B. Mukherjee, M. F. Habib, and F. Dikbiyik, “Network adaptability from disaster disruptions and cascading failures,” *IEEE Communications Magazine*, vol. 52, no. 5, May 2014.

[4] F. Dikbiyik, M. Tornatore, and B. Mukherjee, “Minimizing the risk from disaster failures in optical backbone networks,” *IEEE/OSA Journal of Lightwave Technology,* vol. 32, no. 18, Sept. 2014.

[5] F. Dikbiyik, L. Sahasrabuddhe, M. Tornatore, and B. Mukherjee, “Exploiting excess capacity to improve robustness of WDM mesh networks,” *IEEE/ACM Transactions on Networking*, vol. 20, no. 1, pp. 114-124, 2012.

[6] S. S. Savas, F. Dikbiyik, M. F. Habib, and B. Mukherjee, “Network adaptability to disaster disruptions by exploiting degraded-service tolerance,” *IEEE Communications Magazine*, vol. 52, no.12, Dec. 2014.

[7] A. Roy, M. Farhan Habib, and Biswanath Mukherjee, “Network Adaptability under Resource Crunch: Availability and Revenue Maximization,”*Optical Network Design and Modeling conference* [under review], 2015.

[8] Cisco visual networking index: forecast and methodology, 2011-2016. in White Paper, May 2012.

[9] Department of Homeland Security, “Cloud computing: what are the security implications?” http://www.dhs.gov/news/2011/10/05/testimony-richard-spires-chief-information-officer-house-committee-homeland-security.

[10] M. F. Habib, M. Tornatore, and B. Mukherjee, “Fault-tolerant virtual network mapping to provide content connectivity in optical networks,” *Proc., Optical Fiber Communications (OFC) Conference*, Anaheim, CA, March 2013. (*Semifinalist for Corning Outstanding Student Paper Award; ranked in top 2 out of over 100 submissions in the Core Networking area*.)

[11] M. F. Habib, M. Tornatore, M. De Leenher, F. Dikbiyik, and B. Mukherjee, “Design of disaster-resilient optical datacenter networks,” *IEEE/OSA Journal of Lightwave Technology*, vol. 30, no. 16, pp. 2563-2573, 2012.

[12] S. Ferdousi, F. Dikbiyik, M. F. Habib, M. Tornatore, and B. Mukherjee, "Disaster-Aware Datacenter Placement and Dynamic Content Management in Cloud Networks," *IEEE/OSA Journal of Optical Communications and Networking*, (under review).

[13] T. L. Weems, “How far is far enough,” *Disaster Recovery Journal*, vol. 16, 2003.

[14] online.wsj.com/articles/james-woolsey-and-peter-vincent-pry-the-growing-threat-from-an-emp-attack-1407885281.

[15] S. Ferdousi, M. F. Habib, M. Tornatore, and B. Mukherjee, “Rapid Data Evacuation for Large-Scale Disasters in Optical Cloud Networks,” *Proc., Optical Fiber Communications (OFC) Conference*, Los Angeles, CA, March 2015.

[16] C. Colman-Meixner, M. F. Habib, M. Tornatore, C-N. Chuah, and B. Mukherjee, "Disaster-survivable cloud-network mapping", *Photonic Network Communications*, vol. 27, no 3, pp 141-153, May 2014.

[17] S. S. Savas, F. Dikbiyik, M. F. Habib, and B. Mukherjee, “Disaster-aware service provisioning with manycasting in cloud networks,” *Photonic Network Communications*, vol. 28, no. 2, pp. 123-134, 2014.

DEPARTMENT OF DEFENSE

DEFENSE THREAT REDUCTION
AGENCY
8725 JOHN J. KINGMAN ROAD
STOP 6201
FORT BELVOIR, VA 22060
ATTN: P. TANDY

DEFENSE TECHNICAL
INFORMATION CENTER
8725 JOHN J. KINGMAN ROAD,
SUITE 0944
FT. BELVOIR, VA 22060-6201
ATTN: DTIC/OCA

DEPARTMENT OF DEFENSE CONTRACTORS

QUANTERION SOLUTIONS, INC.
1680 TEXAS STREET, SE
KIRTLAND AFB, NM 87117-5669
ATTN: DTRIAC